The Dissertation Committee for Brian Russell Knab
certifies that this is the approved version of the following dissertation:

Three Problems in Formal Epistemology

Committee:

Sinan Dogramaci, Supervisor

Sahotra Sarkar, Co-Supervisor

Kenny Easwaran

Cory Juhl

Miriam Schoenfield

Three Problems in Formal Epistemology

by

Brian Russell Knab

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2018

To Lindsey and Paisley.

Three Problems in Formal Epistemology

Brian Russell Knab, Ph.D.
The University of Texas at Austin, 2018

Supervisors: Sinan Dogramaci
Sahotra Sarkar

In this dissertation, I offer and defend three theses in formal epistemology: (i) that Bayesianism is consistent with the search for as-yet-unknown explanations of otherwise improbable data; (ii) that there is a viable classical statistical solution to the problem of the priors; and (iii) that the distinction between evidential and causal decision theory is overblown at best, and merely apparent at worst.

Table of Contents

List of Figures

Chapter 1

Origins of Life Research Does Not Rest on a Mistake

1.1 Introduction

Self-replicating molecules are complex. Thus even if each individual step in their natural construction was probable, the overall probability of randomly constructing a complete self-replicator could be minuscule.[2] And given this, some have been unable to rest comfortably with the idea that self-replicators and thus life arose *by chance*; they contend that there must be some further explanation available.[3] Some are right now in laboratories searching for that further explanation.

This chapter is a version of a paper published in 2016. See: Brian Knab. Origins of Life Research Does Not Rest on a Mistake. *Ergo*, 2016.

[2]In order for this to be true, by a 'randomly generated' step, we must mean a step which is probabilistically independent of the prior steps. Also, it is important to note that – granting the random construction of a *particular* self-replicator along a *particular* path was improbable – it does not follow that the random construction of self-replicators in general was improbable. If there were many possible self-replicators, or if, for any particular self replicator, there were many ways to construct that self-replicator, then self-replication would not be an improbable outcome. (Thanks to an anonymous reviewer for this latter point.) As we'll see, I am going to grant that self-replication is, in general, improbable. The question I am concerned with is: granting self-replication is improbable, what is the rational reaction to this improbability?

[3]Roger White attributes this view to origins of life scientists J.D. Bernal, Manfred Eigen, Christian De Duve, and Richard Dawkins.White (2007)

Roger White – in his 2007 paper, "Does Origins of Life Research Rest on a Mistake?" – argues that this search is unmotivated. He does not dispute the claim that it was improbable that life should have arisen by chance. Instead he argues: the fact that it was improbable *is no reason to think that it didn't.*

Here, I defend the searchers against White. His argument overgeneralizes. It entails that it is always irrational to search for an explanation of currently inexplicable empirical data. But this is not always irrational.

Further, I argue that it often *is reasonable* to search for alternative explanations of your data, upon discovering that it is very improbable that that data should have arisen by chance. It is reasonable to do so when you are unsure about the conditions under which your data arose. In recognizing this, we can explain both why we ought to search for an explanation of the origin of life, but also I think, why we ought to search for an explanation in the related case of the fine-tuning of the universe.

1.2 Preliminaries

Allow me some preliminaries. First, I will refer to the hypothesis that an event happened by chance as the *chance hypothesis*. What I mean is: the hypothesis that an event happened by chance, under some (often implicit) dominant probability distribution over the space of events. Suppose that you have two dice and you roll a double six. The *chance hypothesis* (in normal contexts) is the hypothesis that the dice are fair.[4]

[4]The chance hypothesis need not, however, be uniform over the space of atomic events – imagine that we confront a pair of trapezoidal six-sided dice, for example. Here the *dominant distribution* will roughly be the one which assigns a probability to each face equal to the ratio of the surface area of that face to the total surface area of each die. In this

Second, I will assume that a self-replicating molecule, and thus life, was an improbable outcome under the relevant chance hypothesis. De Duve de Duve (1991), for example, claims that the probability of life's arising by chance was less than 1 in 10 raised to the 300th power. Hoyle and Wickramasinghe Hoyle & Wickramasinghe (1981) calculate it to be around 1 in 10 raised to the 40,000th power.[5] Very roughly, this is the probability that self-replicating molecules should arise through the haphazard mixing of the non-self-replicating material present on early Earth. Without getting into the details, it's hard to evaluate these claims. But luckily, we needn't; we're just going to grant them.

Third, the *mere* fact that an event occurred which was improbable, according to some hypothesis, is *not* evidence that that hypothesis is false. I once read that every time you shuffle a deck of cards, and deal a hand in a game of Spades, say, it is likely that that particular hand has never been generated in the entire history of card play. This, because the number of distinct ways to distribute 52 cards among four players is 1.25×10^{28}, which is about 1 billion times greater than the number of grains of sand on Earth.[6] Nevertheless, most hands are not evidence against the hypothesis that the cards were randomly shuffled.

This, however, raises a question: when *is* the occurrence of an event,

---

case, a statement like 'you rolled snake eyes by chance' remains perfectly intelligible, even though the probability distribution over each face is not uniform. Of course there are tricky questions here – why prefer the dominant probability model to others? And how do dominant probability models *become* dominant in the first place? I'm going to try to avoid saying anything here that commits me to an answer to any of these larger questions.

[5]Both of the (de Duve, 1991) and (Hoyle & Wickramasinghe, 1981) references are from White's White (2007) paper, and he attributes them to (Fry, 2000).

[6]According to NPR's estimate of the number of grains of sand on Earth (Krulwich, 2012).

improbable according to some hypothesis, evidence against that hypothesis? The answer, according to White and which we will rely on, is given by the following theorem:

$$P(C|E) < P(C) \quad \text{if and only if} \quad P(E|\neg C) > P(E|C) \tag{1.1}$$

In words: if C is the hypothesis that the cards were randomly shuffled, and we confront some evidence E – i.e., a particular hand – that hand is evidence against the random-shuffling hypothesis if and only if that hand is more likely given the cards were *not* shuffled randomly, than given they were.

In the following, we will not be directly concerned with (1.1), but instead with a necessary condition that follows on (1.1); namely, that if $\{B_1, \ldots B_n\}$ forms a partition of $\neg C$ – that is, if $\neg C$ is equivalent to the disjunction $B_1 \vee B_2 \vee \ldots \vee B_n$, and $B_1, \ldots, B_n$ are mutually exclusive – then

$$P(C|E) < P(C) \quad \text{only if} \tag{1.2}$$

$$P(E|C) < P(E|B_1) \text{ or } P(E|C) < P(E|B_2) \text{ or } \ldots \text{ or } P(E|C) < P(E|B_n)^7 \tag{1.3}$$

Or again in words: Suppose I know that the cards were either shuffled randomly *or* that either Bob stacked the deck or Bill did. Then I get evidence the cards were *not* shuffled randomly only if my hand is more likely given Bob stacked the deck, or given Bill did.

---

[7] Suppose that for all i, $P(E|C) \geq P(E|B_i)$. Then we have

$$P(E|\neg C) = \sum_i P(B_i \& E|\neg C) = \sum_i P(E|B_i \& \neg C) \cdot P(B_i|\neg C) = \sum_i P(E|B_i) \cdot P(B_i|\neg C)$$
$$\leq \sum_i P(E|C) \cdot P(B_i|\neg C) = P(E|C) \cdot \sum_i P(B_i|\neg C) = P(E|C) \cdot P(\neg C|\neg C) = P(E|C)$$

Hence by theorem (1), $P(C|E) \geq P(C)$

1.3 White's Example

Preliminaries discharged, I turn now to White's argument; it begins with an example.

Suppose we are headed to the English seaside. And consider three things we might confront upon arrival:

Pebble-R The pebbles on the beach are scattered haphazardly over the beach.

Pebble-N The pebbles "cover the beach in descending order of size toward the shoreline."

Pebble-I The pebbles are "arranged to form a stick figure with a smile on its face" (White, 2007, 455).[8]

Each of these patterns could arise by chance – that is, via the standard physical processes by which pebbles come to lie on beaches, whatever they are. And let us suppose that each of our three patterns is equally likely under this chance hypothesis. (Nothing turns on this.)

White notices that if we confronted *Pebble-N* or *Pebble-I*, we would be suspicious of the chance hypothesis – we would not think that the pebbles just happened to be ordered by size, or arranged into the shape of a stick figure, by a mechanism randomly tossing pebbles onto the beach.

And according to (1.1), we get evidence that a pebble arrangement did not arise by chance only if it was more likely given it did *not* arise by chance than given it did.

[8] *Pebble-R* for 'random', *Pebble-N* for 'non-intentionally biased' and *Pebble-I* for 'intentionally biased.'

Now, for White, to say that an arrangement did not arise by chance is just to say that the process which generated it was *biased* in some way. And pebbles arranged in order of their size or into the shape of a stick figure helpfully exemplify the two distinctive ways White thinks a process might be biased. He writes,

> A process such as pebble arranging is *intentionally biased* if certain elementary possible outcomes are more likely than others due to the purposeful action of some agent... A process is *non-intentionally biased* if this biasing is [due to]..., say, the impersonal laws of nature together with properties of matter and the structure of physical mechanisms (White, 2007, 462).

Call our evidence E and the chance hypothesis C. Let I be the hypothesis of intentional biasing, and NI be the hypothesis of non-intentional biasing. Then, by the necessary condition specified in (1.3) above, we have

$$P(C|E) < P(C) \text{ only if } P(E|I) > P(E|C) \text{ or } P(E|NI) > P(E|C) \quad (1.4)$$

That is, the chance hypothesis is disconfirmed only if our data was more likely given intentional or non-intentional biasing, than given chance.

Note that were we to confront pebbles arranged in order of their size or into the shape of a stick figure, this necessary condition would be satisfied. White writes about the former,

> That [the pebbles] should be arranged roughly in order of size seems more likely on the assumption that the process shuffling the pebbles was non-intentionally biased, than if they just fell on the beach by chance... [If] attributes of the pebbles interact with the physical laws ... simple correlations between physical parameters such as size and location are the kind of phenomena we should expect to find (White, 2007, 462).

And about the latter,

> The stick figure is one of a small class of interesting patterns, and if an agent was to go to the trouble of influencing the way the stones are arranged, there is a good chance she would arrange them in an interesting way (White, 2007, 462).

White further notices, however, that if we confronted pebbles haphazardly scattered across the beach, (1.4) would *not* be satisfied. Given an agent arranged the pebbles, the particular haphazard scatter we witness would not be likely. Further there aren't any obvious correlations between the physical attributes of the pebbles, so neither is the arrangement likely given non-intentional biasing. Intentional and non-intentional biasing thus make the haphazard arrangement just as unlikely as the chance hypothesis. And therefore a haphazard scatter does not call the chance hypothesis into question.

I think that this is an illuminating answer to the question of why certain arrangements of pebbles at the English seaside might cause us to suspect the chance hypothesis, while other, perhaps equally unlikely, arrangements under that hypothesis would not.

White now wants to generalize to the origins of life. He wonders: are molecules – arranged into self-replicating structures – akin to pebbles scattered haphazardly across a beach, or are they instead like pebbles arranged in order of their size, or perhaps, like pebbles arranged into the shape of a stick figure?

1.4 White's Argument

The rest of White's argument is now straightforward: First, he notes that *intentional biasing* in favor of life is not a hypothesis taken seriously by

most scientists. Thus our evidence calls the chance hypothesis into question only if life was more likely given *non-intentional* biasing than it was given chance. But, White argues, life was *not* more likely given non-intentional biasing than given chance. And hence we have no reason to doubt the chance hypothesis.

That argument is valid, and I won't dispute the claim that scientists do not take the *intentional biasing* hypothesis seriously. But why should we agree with White that – given the process was *non-intentionally biased* – life would not be a likely outcome, or at least more likely than it would be given mere chance? White writes,

> I can't imagine why anyone would think [that life is more likely, given the process by which it arose was non-intentionally biased]. While there is at least room to argue that a *rational agent* is likely to influence [the process] in order to allow for the evolution of life, to suppose that *impersonal physical laws* are likely to constrain [it] in this way can only be based on a confused anthropomorphism ... Even if the value we attach to life is something objective, whether it be moral or aesthetic, or whatever, it could only conceivably have influence on the behavior of an *agent.* Blind physical laws are no more naturally drawn toward states of affairs with value than blind chance is (White, 2007, 466).

And later,

> Are self-replicating, life producing molecules more likely to appear on [the assumption that the process by which life arose was non-intentionally biased]? ... What makes certain molecular configurations stand out from the multitude of possibilities seems to be that they are capable of developing into something which strikes us as rather marvelous, namely a world of living creatures. But

> there is no conceivable reason that blind forces of nature or physical attributes should be biased toward the marvelous (White, 2007, 467).

In conclusion, White writes, "unless we suspect that life arose on purpose, we should be quite content . . . in seeing life as an extremely improbable, 'happy accident"' (White, 2007, 467).

This, then, is the mistake that origins of life research rests upon: scientists, searching in their laboratories, mistakenly believe that they have good reason to suppose the chance hypothesis is false, and hence that some other non-chancy explanation is available. But their data does not disconfirm the chance hypothesis, and hence they have no reason to search for an alternative explanation.

1.5 Counter

Despite White's argument, I think the searchers are justified in seeking an alternative explanation of life's origin; I turn now to their defense.

First, consider a case. Suppose we filled 1,000 paper bags full of the molecules thought to be plentiful on early Earth, and we shook them up. And suppose that when we dumped out the bags, 90 percent of them yielded self-replicators. It *could happen* that 90 percent of our bags yield self-replicators even if no bias exists, and thus the probability of generating a self-replicator is minuscule. Nevertheless, I contend, it would not be reasonable to believe this chance hypothesis in the face of our data. We should not be content, in other words, to see our paper bag self-replicators as an "extremely improbable, 'happy accident.'" And this would be true, I think, even if (i) we are

unwilling to countenance intentional intervention as a plausible explanation of the outcome, and (ii) we cannot at present tell any plausible story which would explain why "the impersonal laws of nature together with properties of physical matter and the structure of physical mechanisms" should favor self-replicators.

For a more realistic case, but with a similar upshot: suppose that we discovered that the rate of heart disease in San Diego was twenty times higher than in other cities of similar size and demographics. That would be strong evidence that something in San Diego is amplifying (or something in the other similar cities is suppressing) heart disease. And again it would not be reasonable, in the face of that data, to conclude that the amplified rate of heart disease in San Diego is just a sad accident – that the rate of heart disease in San Diego just, by chance, happens to be twenty times higher than in similar cities. And this would be true even though (i) agential intervention is not a plausible explanation of amplified rates of heart disease in San Diego, and (ii) we have no antecedent reason to believe that blind physical laws and forces of nature have it out for San Diegans.

So, data can call a chance hypothesis into question, or render it unreasonable to believe, even when (i) agential intervention is not a plausible explanation of that data, and (ii) we cannot presently tell a story about why blind physical laws and physical forces should favor that data. And thus, the fact that we cannot presently offer "any conceivable reason that blind physical forces of nature or physical attributes should be biased toward the marvelous" does not entail that the existence of the marvelous – i.e., the presence of life on Earth – fails to call into question the chance hypothesis, nor does it require that we view the marvelous as a happy accident. This is true even if, as with

our bags of molecules and our San Diegans, we're committed to the view that *intentional biasing* is not a plausible explanation of the outcome.

1.6 Diagnosis

But how can this be? Didn't White rely on a *theorem* of the probability calculus? Am I denying math or Bayesianism or both?

White did rely on a theorem. And I am neither denying Bayesianism, nor math. Here is the theorem again: Where $\{B_1, \ldots, B_n\}$ forms a partition of $\neg C$,

$$P(C|E) < P(C) \quad \text{only if} \tag{1.2}$$

$$P(E|C) < P(E|B_1) \text{ or } P(E|C) < P(E|B_2) \text{ or } \ldots \text{ or } P(E|C) < P(E|B_n) \tag{1.3}$$

To see why what I've said is consistent with this theorem, consider another example: suppose that a coin has been discovered, deep in a tectonic fissure, clearly formed by blind physical laws and forces of nature. And now suppose that I hand you a coin, and I tell you: this is either a standard, fair coin, or it's the tectonic coin. (The tectonic coin is, of course, indistinguishable from a standard coin.) We'll flip it, and see how our subjective probability functions evolve.

We have no reason to think that blind physical laws and forces of nature should favor heads over tails or vice versa, and so, let's suppose, if we knew the coin was tectonic, we would adopt a uniform distribution over every possible weighting.[9]

[9]Setting Bertrand's paradox aside at our peril.

Now suppose we flip the coin twice, and it comes up heads both times. Notice that the chance hypothesis *is disconfirmed* by this evidence.[10] And because it's a theorem, you must therefore satisfy the necessary condition (1.3). In the case of the tectonic coin you do satisfy (1.3) simply because double heads is *more likely* given the coin is tectonic than it is given it's fair.

But the moral is: you can think that there are only two alternatives – a chance hypothesis versus a blind physical process that is just as likely to be biased against, as in favor of, each particular elementary outcome – and yet in witnessing those elementary outcomes receive evidence that disconfirms the chance hypothesis.

Perhaps, returning to our earlier examples, when we started shaking our bags of molecules, we thought that blind physical processes were just as likely to favor as to disfavor self-replicators. Nevertheless, if the chance hypothesis is that each step in a molecular construction occurs randomly, witnessing enough self-replicators *will* disconfirm that hypothesis. The same goes for our San Diegans; perhaps prior to witnessing any data, we thought blind physical forces were just as likely to suppress as to amplify the rate of heart disease in San Diego. Nevertheless, if the chance hypothesis is that, in fact, nothing is amplifying or suppressing heart disease in San Diego, witnessing enough heart disease will again disconfirm that hypothesis.

---

[10]

$$P(HH|\text{Fair}) = \frac{1}{4}$$

$$P(HH|\text{Tectonic}) = \int_0^1 P(HH|w) \cdot P(w)dw = \int_0^1 w^2 \cdot 1dw = \frac{1}{3}$$

Hence by theorem (1.1), Fair is disconfirmed.

The illicit inference that White draws is: because we have no reason to think that a blind physical mechanism would favor the marvelous over the non-marvelous, that therefore our *total marvelous evidence* was just as likely to be produced by chance as by that blind physical mechanism. The latter doesn't follow from the former.

1.7 Responses

I see two possible responses available to White. First, there is something a bit strange about our tectonic coin, which I have passed over thus far in silence. Perhaps you noticed that had we flipped our coin only *once*, then the chance hypothesis would *not* have been disconfirmed by a heads outcome.[11] And so White might contend: our *actual* evidence, in the case of the origins of life, is not like double heads, but like a single heads.

But first, to stave off any confusion, it is worth pointing out that sometimes a chance hypothesis can be disconfirmed by a single data point. Suppose that I hand you a coin which is either weighted to come up tails with probability .75, *or* it is the tectonic coin. Suppose we flip the coin once, and it comes up heads. If the chance hypothesis is that the coin is the tails-weighted coin, then that hypothesis is disconfirmed by a single heads outcome.

Second, White could perhaps contend that our evidence is like a single

[11]

$$P(H|\text{Fair}) = \frac{1}{2}$$
$$P(H|\text{Tectonic}) = \int_0^1 w \quad dw = \frac{1}{2}$$

heads outcome, simply in that it is equally likely under the chance and non-intentional bias hypotheses. But my point is that he must *argue* for this conclusion. White *infers* from the fact that "there is no conceivable reason that blind forces of nature or physical attributes should be biased toward the marvelous" that therefore our evidence is equally likely under the chance and non-intentional bias hypotheses. But, again, this doesn't follow. We can suppose the chance and the blind physical forces hypotheses *do* assign equal probability to the generation of a single self-replicator. Nevertheless, if our data has a certain composition – if it's akin to a tectonic coin coming up heads twice – then the chance hypothesis will yet be disconfirmed by our data.

So in order for White to successfully argue that the chance hypothesis is not disconfirmed, he has to (i) tell us what the chance hypothesis is, (ii) *argue* that it assigns to self-replicators a probability equal to or greater than the probability assigned by the hypothesis of unknown bias, and finally (iii) argue that our data is not composed in such a way that it nevertheless disconfirms the chance hypothesis. And that strikes me as a tall order.

The second response available to White is this: he can argue that while perhaps he hasn't definitively shown that our data fails to disconfirm the chance hypothesis, neither have origins of life researchers definitively shown that it *does* disconfirm the chance hypothesis.

Allow me, then, to offer a toy model to the origins of life researchers, which I think can plausibly be extended to a probabilistic model of the origins of life, and according to which a self-replicator *does* disconfirm the chance hypothesis.

Imagine Nature sitting down at a 26-letter typewriter, and think of every key on the typewriter as a molecule present on early Earth. *A priori*, it

seems that Nature is just as likely to be biased in favor of typing the letter a as she is to be biased in favor of typing any letter, let's suppose, and this induces a uniform prior distribution over the space of possible bias hypotheses.[12] Let us suppose further that the chance hypothesis is that every key is equally likely to be struck. Nature then begins to type.

Certain strings, of course, would do nothing to disconfirm the chance hypothesis, e.g.,

$$alkwersbn^{13}$$

But certain other fairly mundane strings *would* disconfirm the chance hypothesis, e.g.,

$$alkklaaal^{14}$$

---

[12] Officially, this will induce a Dirichlet prior with 25 concentration parameters all equal to, say, 1.

[13]

$$\begin{aligned} P(alkwersbn|Chance) &= \frac{1}{26^9} \approx 1.8 \times 10^{-13} \\ P(alkwersbn|\text{Unknown Bias}) &= \int_W P(alkwersbn\&W)dW \\ &= P(alkwersbn|W) \cdot P(W)dW \\ &= \int_W w_a w_l w_k w_w w_e w_r w_s w_b w_n \cdot \Gamma(26)dW \\ &= \frac{\Gamma(26)}{\Gamma(35)} \approx 5 \times 10^{-14} \end{aligned}$$

Hence the chance hypothesis is confirmed.

[14]

$$\begin{aligned} P(alklaaal|Chance) &= \frac{1}{26^9} \approx 1.8 \times 10^{-13} \\ P(alklaaal|\text{Unknown Bias}) &= \int_W w_a^4 w_k^2 w_l^3 \cdot \Gamma(26)dW \\ &= \frac{\Gamma(26) \cdot \Gamma(5) \cdot \Gamma(4) \cdot \Gamma(3)}{\Gamma(35)} \approx 1.5 \times 10^{-11} \end{aligned}$$

Hence the chance hypothesis is disconfirmed.

What is the difference between *alkwersbn* and *alkklaaal*? Repetition. Repetition is indicative of bias in favor of the repeated outcomes, and is also evidence that not every outcome is equally likely; i.e., it is evidence that the chance hypothesis is false.

Thus, to complete the argument, we can simply note that, because self - replicating molecules consist of repeated molecular substructures, those molecules disconfirm the chance hypothesis. That, in other words, is a reason to think that self-replicating molecules are more likely to result via blind forces of nature than via chance, and it is a reason which simply rests on a prior uniform distribution over the space of bias hypotheses, and not on a confused anthropomorphism.

Now, to be fair, White does anticipate something like the story I am telling here, and he writes in response that, in fact, "the picture given by de Duve and others is [that] ... the molecular parts required to make up the replication machinery come in various sizes and structures, and they are not arranged in anything like a simple repetitive pattern" (473). In other words, self-replicating molecules are more like the string *alkwersbn* than they are like the string *alkklaaal*.

But that strikes me as frankly incredible, and White does not give us a specific reference. You cannot look at the double helix of DNA, which is a (two-meter long, in humans,) molecule consisting of a sugar, a phosphate, and four nucleobases repeated over and over and over, and not see a simple repetitive pattern. It is as though Nature sat down at the typewriter described

above, and typed

$$AAAGTCTGACAAGCTACGCGGG\ldots$$ [15]

And *that*, for the same reason as the string *alkklaaal*, will certainly disconfirm the chance hypothesis.

Of course, this model is not a perfect probabilistic representation of the origins of life. But that's not the point. The point is only that its general contours look right. And on any model which shares those general contours, a self-replicating molecule will disconfirm the chance hypothesis.

1.8 Fine Tuning

We have thus far been considering the molecular origins of life; I now want to turn to another case White discusses that is related, and which is I think more amenable to his argument: the apparent fine-tuning of the universe.

Some have noticed, for example, that had the rate of expansion of the early universe been slightly different, then a stable universe capable of producing and sustaining life would not have existed.[16] They have inferred from this that it was extremely improbable that a universe like ours should exist. And here, we have access to only a *single* data point – i.e., a single universe. Thus we cannot rely, as in the case of the molecular origins of life, on the presence of repetition to show that the chance hypothesis is disconfirmed.

[15] A, G, T, and C, of course, for the nucleobases adenine, guanine, thymine, and cytosine, which compose the nucelotides that compose DNA.

[16] See (Hawking, 1996, 156), "If the rate of expansion one second after the Big Bang had been smaller by even one part in one hundred thousand million million, the universe would have re-collapsed before it ever reached its present state."

Though that last paragraph does not establish it, let us grant that, under the chance hypothesis, it was unlikely that our life-supporting universe should exist. If we're unwilling to countenance agential intervention as an explanation, must we, by White's argument, accept that a fine-tuned universe was an extremely improbable 'happy accident'?

The answer is 'no,' and this brings me to a second, much simpler, objection I want to raise to White's overall argument.

Suppose White is right that our evidence does not *disconfirm* the hypothesis that life and the universe arose by chance. And consider one last case. Suppose that in addition to a coin, our tectonic fissure had yielded a die with, say, one quintillion sides. We've analyzed the die in the lab, and have discovered that the die is either fair, or it is weighted to only come up 1 or to only come up 2 or ... or to only come up 1 quintillion. Along with White, we dub the hypothesis that the die is fair the 'chance hypothesis.'

Because we have no reason to favor any one of these hypotheses, it seems reasonable to adopt a uniform distribution over each weighting hypothesis. That is, given what I've said so far, we should think it equally like that the die is fair as that the die is weighted to only come up in one particular way. Now, suppose we roll the die only once, and it comes up, say, 42. What should we think? Should we think that 42 was just an extremely improbable 'happy accident'?

We should not! If we're Bayesians, upon witnessing 42, we should think it *1 quintillion times more likely that the die is weighted to come up only 42,*

than that it is fair.[17]

Put simply: the fact that the chance hypothesis is not *disconfirmed* by our evidence does not entail that we should think the chance hypothesis is true, and that therefore the outcome was a happy accident. The important question, if we're Bayesians, is not 'was the chance hypothesis disconfirmed?' but instead 'is the chance hypothesis, in light of our data, at all likely to be true?'

I think that our epistemic situation, with respect to the origins of life and the universe is directly analogous to our epistemic situation with respect to our quintillion-sided tectonic die. (Think of each face of the die as a possible rate of expansion of the early universe, or as a possible molecular configuration.) While it's perhaps possible that a fine-tuned universe, or a self-replicating molecule, was a very unlikely chancy accident, it is *also* possible that a fine-tuned universe and a self-replicating molecule were very likely or inevitable. And given our limited access, in both cases, to the conditions under which our data arose, it seems to me we ought to assign these hypotheses roughly equal prior weight. Hence, upon conditionalizing on our data, the hypothesis we ought to now think most likely was that life and the universe

[17]

$$
\begin{aligned}
P(\text{Fair}|42) &= \frac{P(42|\text{fair}) \cdot P(Fair)}{P(42|Fair) \cdot P(\text{Fair}) + P(42|\text{Only 42}) \cdot P(\text{Only 42})} \\
&= \frac{(1/1 \text{ quintillion}) \cdot (1/(1 \text{ quintillion} + 1))}{(1/1 \text{ quintillion}) \cdot (1/(1 \text{ quintillion} + 1)) + 1 \cdot (1/(1 \text{ quintillion} + 1))} \\
&= \frac{1}{1 \text{ quintillion} + 1} \\
\Rightarrow P(\text{Only 42}|42) &= \frac{1 \text{ quintillion}}{1 \text{ quintillion} + 1}
\end{aligned}
$$

were very likely or inevitable. And, given that, it seems to me we ought to head into the laboratory, to try to figure out exactly why that should be.

All of this, recall, is presuming that agential intervention is off the table – that we're saddled with the prejudice that the universe and our tectonic die are the result of blind physical forces. Blind physical forces, the point is, are just as likely to produce a fair die as they are to produce one loaded in some particular way.

There is a more sophisticated way of putting this point, which I will rehearse for the aficionados in the audience: Every version of the fine tuning argument that I have seen assumes the standard measure over the parameter space. The standard measure is roughly the continuous equivalent of a uniform distribution over the faces of our die. My point is that our scant evidence about the origins of the universe cannot justify high confidence that the standard measure accurately describes the parameter space. For *a priori* – that is, in the absence of any data – it seems equally likely that the parameter space is described by a measure which makes the observed parameter value likely or inevitable, and the latter hypothesis is massively confirmed by a single data point. Now, if we had access to 1-billion universes, and if their observed parameter values were not clustered in any obvious way – if they seemed, in some sense, uniformly drawn from the Real line – that *would* I think justify some confidence that the parameter space is accurately described by the standard measure. But we do not have access to 1-billion universes. A full defense of this point, however, will have to wait for another time. For now it's enough to point out that White's conclusion – that we should be content to see life and the universe as a happy accident – only follows on the strong claim that we should be antecedently confident that the standard measure is the true

measure on the parameter space, or that the chance hypothesis is true, and White has given us no reason to believe that.

Here then, is a cautionary note: often when we're thinking about chance hypotheses and Bayesianism, we rely heavily on standard examples of chance mechanisms: dice, cards, urns, roulette wheels, lotteries, tornadoes, etc. But the problem with the use of these examples is that, when you confront them in a thought experiment, you do so with certain presuppositions. If I asked you to imagine a die, for example, and asked how many times in a row it would have to come up six before you'd be convinced it was loaded, the answer I imagine would be fairly large. Certainly, if the die came up six *once*, that wouldn't be enough. But all this reveals is that you presuppose, in the thought experiment, that the chance hypothesis – that the die is fair – is true. This is not unreasonable, of course, because you've presumably had a lot of experience with dice, and dice-like physical objects, and those experiences justify confidence in how those objects will behave. But that is *not* the case when we consider the molecular origins of life, and it is certainly not the case when we consider the conditions under which the universe arose. We do not have any evidence or experience which would justify an antecedently high confidence that a certain chance explanation accurately describes these processes. Given that, the standard examples of chance mechanisms are unhelpful as analogies, and are in fact distorting. We should not respond to a self-replicating molecule, or to a fine-tuned universe, the same way we should respond to the outcome of a single roll of a die, or a single spin of a roulette wheel, etc.[18]

I think that White himself makes this mistake. He imagines, as analogous to the origins of life and the universe, confronting a massive lottery,

[18]This point was made clear to me by Sahotra Sarkar, in conversation.

which selects ticket #48579387593478. And he writes,

> No matter how confident or doubtful we are initially that the lottery is fair in that tickets are selected purely by chance, ticket #48579387593478's being selected gives us no reason at all to doubt this. Any investigation into the lottery mechanism that was motivated by the fact that ticket #48579387593478 was selected would be misguided. (468)

That sounds plausible, but I contend its plausibility derives precisely from the distortion described above. If you are a Bayesian, and prior to witnessing any data, you took the chance hypothesis to be just as likely as the hypothesis that the lottery is biased in favor of any particular ticket, then you should now think it overwhelmingly likely that the lottery was biased in favor of ticket #48579387593478. And thus an investigation into the lottery mechanism – in an attempt to explain why the lottery is biased in favor of ticket #48579387593478 – would not be misguided.

1.9 Summary

Allow me to sum up White's argument, and my case against it, before considering objections.

White argues that because blind forces of nature are as likely to be biased in favor of life as they are to be biased in favor of any particular outcome, our data is equally unlikely under the hypothesis that it resulted from chance, and under the hypothesis that it resulted from those blind physical forces. It follows from this that our data does not disconfirm the chance hypothesis. White concludes that therefore "we should be quite content to ... see[] life as an extremely improbable 'happy accident.' "

I have raised two objections to this argument. First, I have argued that it is not at all clear that our data is equally likely under the chance and blind physical forces hypotheses. We can only know this if we specify a chance hypothesis, and then argue that it assigns a lesser probability to our data than the hypothesis of unknown bias. Importantly, it does not follow simply from the fact that the hypothesis of unknown bias *is* the hypothesis of unknown bias that therefore it assigns the same probability to our data as the chance hypothesis. Furthermore, I've offered what I think is a plausible toy probabilistic model of the origins of life, according to which our data would *not* be equally likely under the chance hypothesis, and the hypothesis of unknown bias.

Second, I have argued that, even if we *grant* that the chance hypothesis is not disconfirmed by our data, it simply does not follow that we should think, upon witnessing our data, that that hypothesis is likely to be true, nor does it follow that we should be 'quite content' to endorse it. While our data may not give us any reason to *doubt* the chance hypothesis, it may nevertheless give us a (very) strong reason to believe some alternative to the chance hypothesis is true. This alternative could simply be, for example, that our data were likely or inevitable. And if, antecedently, we thought the chance and the 'it was inevitable' hypotheses roughly equally likely – as it seems we should if we are uncertain about the conditions under which our data arose – then we should now take the most likely hypothesis to be that our data were inevitable. The task of the scientist, then, would be to explain *why* it is that our data – complex molecules, or cosmic parameters, or amplified rates of heart disease – was inevitable.

1.10 Objections and Replies

1.10.1 Objection 1

White argues that, in the absence of an available explanation of our data we should be content to endorse the hypothesis that the data improbably arose by chance. But in your reply to White, you rely heavily on the hypothesis 'the data were likely or inevitable,' which itself looks like a sort of explanation. So isn't your reply irrelevant, because White is explicit in his argument that no alternative explanations are available?[19]

1.10.1.1 Reply

It is true that I rely on the hypothesis 'the data were likely or inevitable,' but I do not think this a problem. Were White to pursue this objection, he would face a dilemma. If his argument only applies when the hypothesis 'the data were likely or inevitable' is not an available hypothesis, then his argument is irrelevant with respect to the question of the origins of life and of the universe, because it clearly *is* possible that life (and the universe) were likely or inevitable. So White would either have to admit that that alternative is available – in which case he confronts my reply – or he has to contend that his argument is purely academic, and in fact irrelevant to the interesting question of how we should respond to the fact that life and the universe are improbable outcomes under the chance hypothesis.

[19]Thanks to a comment from an anonymous reviewer which inspired this objection.

1.10.2 Objection 2

You are reading White uncharitably. White's only goal was to argue that, when considering the origins of life, life's improbability is not, *in itself*, a reason to doubt the chance hypothesis. Thus it is consistent with White's argument that the chance hypothesis *is disconfirmed* by life's existence, as long as it is disconfirmed for a reason which is not the one White is objecting to. Thus the first objection – the conclusion of which is that, in the case of the origins of life, the chance hypothesis is disconfirmed – is consistent with White's argument. Furthermore, it is also perfectly consistent with White's argument that the chance hypothesis, in both your prior and posterior, is extremely unlikely. So the second objection offered in the above misses the mark as well.[20]

1.10.2.1 Reply

I do take White to be offering a positive argument that *in fact* the chance hypothesis is not disconfirmed by our data. One bit of circumstantial evidence is simply the length of his paper; that improbability does not *entail* disconfirmation is a point which could easily be made in a page or two. (I make it in a short paragraph in section 2.) But more than that, White seems to me quite explicit about this in his paper; upon finishing his argument, he writes:

> Where does this leave us? If life's existence is no more to be expected on the assumptions of either intentional or non-intentional biasing than it is on chance, then we have no reason to doubt the

[20] Thanks to an anonymous reviewer for this objection.

> Chance hypothesis. I have been arguing that while there is at least room to argue that life is more to be expected given that an agent was involved, it is very hard to see why we should find life's existence any more likely at all on the assumption that non-intentional biasing factors were involved. So unless we suspect that life arose on purpose, we should be quite content to join Crick in seeing life as an extremely improbable 'happy accident' (White, 2007, 467).

Now, we might try to read White as *merely* arguing that the chance hypothesis is not disconfirmed by the data. In which case, while my first objection would be on target, my second objection – that, granting the chance hypothesis is not disconfirmed, the chance hypothesis could nevertheless be very unlikely in your posterior – would miss the mark, because it would be consistent with White's overall argument.

But again, I think there is pretty compelling evidence that White wants to conclude something stronger – namely, that in light of the fact that the chance hypothesis is not disconfirmed, we should think it plausibly true. White's motivating question is "why, if appeals to intelligent agency are not on the table, we should be so reluctant to attribute the origin of life largely to chance?" (White, 2007, 454). He wonders why "the vast majority of researchers in the field agree with Dawkins that we cannot credibly suppose that life arose by spontaneous random generation if the chance of this happening was extremely small?" (White, 2007, 460). He takes the upshot of his argument, as just noted, to be that we should be "quite content to join Crick in seeing life as an extremely improbable "happy accident"" (White, 2007, 467). He claims that the reasoning of researchers is "misguided" because "while making no appeal to intentional agency, [they] are persuaded that [life] was not the result of chance, and are motivated to find a non-intentional explana-

tion." (White, 2007, 470) He claims that he will "raise doubts" about how the following two claims can "hang together" (White, 2007, 453):

(3) "The conviction that life did not arise largely by chance is treated as *epistemically prior* to the development of alternative theories" (White, 2007, 453).

(4) "The suggestion that the origin of life might be due to any kind of purposeful agency is not considered as a serious option" (White, 2007, 454).

My point – that given our lack of access to the conditions under which our data arose, chance and bias hypotheses could and probably should be on equal footing in your priors, and hence should be given very little prior weight – answers the question of how these last two claims can hang together, even if the chance hypothesis is not disconfirmed by the data. Because it is reasonable to assign a low prior to the chance hypothesis, you should now be convinced, in your posterior, that life did not arise by chance. This would explain why the vast majority of researchers agree with Dawkins, why they aren't content to join Crick, why they are persuaded that life was not the result of chance, and why they are motivated to find a non-intentional explanation.

Finally, even if my reading of White is incorrect, and his intention was only to point out that improbability does not suffice for disconfirmation, or to argue that the chance hypothesis is not disconfirmed by the data, there would remain an interesting question: granting the chance hypothesis is not disconfirmed by their data, should scientists be motivated to search for an alternative-to-chance explanation of the origins of life and the universe? And here I've given an answer: yes, under plausible probabilistic models of their

epistemic situation, and even if the chance hypothesis is not disconfirmed, the hypothesis that researchers should now think most likely is that their data were likely or inevitable, and hence they should be motivated to account for why that should be.

1.10.3 Objection 3

In the discussion of the tectonic coin of unknown bias, you point out that double heads would disconfirm the chance hypothesis, because double heads is more likely if the coin is tectonic than it is if the coin is fair. But White argues *directly*, in the case of the origins of life, that our evidence is equally likely under both chance and non-intentional bias hypotheses. Hence the case of the tectonic coin – which produces two heads – is not relevant to White's argument, because it's not a case where the evidence is equally likely under chance and bias hypotheses.[21]

1.10.3.1 Reply

White's *reason* for thinking that our evidence is equally likely under the chance and non-intentional bias hypotheses is that blind physical forces are just as likely to be biased in favor of life as they are to be biased in favor of any outcome. This is *also* true of the tectonic coin – it is equally likely to be biased in favor of as against heads. So the tectonic coin is relevant because it reveals that you can observe a random process, which is as likely to be biased in favor of as against any particular elementary outcome, and yet receive evidence that is *not* equally likely under the chance and bias hypotheses. Thus my

[21]Thanks again to an anonymous reviewer for this objection.

point, with the tectonic coin, is simply that White draws an illicit inference: it does not follow from the fact that non-intentional bias is as likely to favor life as to disfavor it that our evidence is equally likely under the chance and non-intentional bias hypotheses. More is required.

Recall, also, our 1,000 paper bags full of the non-self-replicating stuff of early Earth. Suppose upon shaking them up, every single one yielded a self replicator. Suppose 1 million or 1 billion such bags produced self replicators. Or consider our San Diegans again; suppose the rate of heart disease in San Diego was 100 percent – that is, every San Diegan had heart disease. If White's inference here were successful – that is, if it followed that because blind physical forces are as likely to be biased in favor of, as against, an outcome, that therefore *whatever* evidence we receive is equally likely under the chance and non-intentional bias hypotheses – then *no* data stream could call the chance hypothesis into question (unless we are willing to attribute that data stream to an intentional agent).

That strikes me as absurd. I am right now fairly confident that heart disease in San Diego is distributed via the same chance mechanisms which distribute heart disease in the general population. But if tomorrow I learned that the rate of heart disease in San Diego was 100 percent, I would not retain that confidence. And importantly we can build plausible Bayesian models of the idea that blind physical forces are, antecedently, as likely to be biased in favor as against heart disease in San Diego which impose no such commitment.

Chapter 2

Probably Not that Improbable: On inverse probabilities and the problem of the priors

2.1 Introduction

You go to the doctor. She tests you for a disease. The true positive and negative rates of the test you undergo are both 95 percent; i.e.,[1]

$$P(+|D) = .95$$
$$P(-|\neg D) = .95.$$

You test positive. How confident should you be that you are afflicted?

My students would respond that there is not enough information. To answer the question, they would need to know the base rate. If the disease is rare, then it's unlikely you are afflicted. But if it's common, then you should worry.

But here is a puzzle. What if you, the doctor, and everyone else, have no idea what is the base rate? What if your ignorance is so acute that you cannot place non-trivial upper and lower bounds on what it might be? It is possible everyone has the disease, or that no one does. In fact, for every rate r, it's possible the base rate is r. What then is the rational response to a

[1]Here 'D' is the proposition you have the disease, '+' the proposition that you test positive, and '-' the proposition that you test negative.

positive result? This question might seem obtuse, because we are never so entirely ignorant. But first that's not so clear. And second the question I'm really asking is 'how does empirical inquiry *begin*, given that it must ultimately begin from a state of empirical ignorance?'

The entirety of statistics and scientific inference hangs on this question. The specific example is not important. I have in mind an idealized picture of the scientific enterprise. We confront a universe of random mechanisms. Those mechanisms distribute disease, or inherited traits, or economic gains and losses. And our goal is to understand the probability distributions which govern them. But there is always a first step: a first piece of data, a first positive result. And it is only if we know how to take the first step that we will be able to take the rest of them.

This, then, is a paper in the foundations of statistics. It is a defense of a classical outlook much like the one defended by R.A. Fisher in the middle of the last century. But I am not obsequious; I argue here that Fisher got a lot wrong.

In the philosophical circles I run in, Bayesianism is ascendant. But while the Bayesian apparatus is elegant and powerful, I can't make sense of it. This, for familiar reasons: Bertrand's paradox, and the problem of the priors more generally. The other main alternative – besides Bayesianism and classicism, that is – is likelihoodism. But, though the central claim of likelihoodism is right, likelihoodism is too austere. It seems to me unable to do the work that I had hoped statistics could do.

This paper proceeds in two parts. The first is destructive. I argue that each of the views mentioned fails to solve our motivating puzzle. The second is constructive; it is devoted to solving it.

2.2 Destruction

2.2.1 Bayesianism

Everyone wants to be a Bayesian. For Bayesians, uncertainty always manifests itself in a probability distribution. If you are ignorant of the base rate, there is a probability distribution which expresses or captures that ignorance. But the difficult question is: which distribution?

In answering that question, Bayes considered, and Laplace championed, the Principle of Indifference: evidential parity implies equiprobability. If you are ignorant of the base rate, then you have as much evidence that you are afflicted as that you are not. Hence, before taking the test, the probability that you had the disease must have been 1/2.

But there are damning, perennial, recalcitrant objections to that thought. Suppose that the disease you will be tested for is one of a set of one hundred mutually exclusive diseases. And you are ignorant of each of their base rates. Then for any disease, you have as much evidence you are afflicted as that you are not. (You have exactly zero evidence, either way.) Thus, if evidential parity implies equiprobability, the probability you have each particular disease must be 1/2. Hence the probability that you have at least one of the diseases is given by

$$\begin{aligned} &P(\text{You have the first disease}) + P(\text{You have the second}) + \ldots \\ &\qquad + P(\text{You have the hundredth}) \\ &= 1/2 + 1/2 + \ldots + 1/2 \\ &= 50 \end{aligned}$$

But 50 is not a probability. Probabilities are less than one. The Principle of Indifference entails a contradiction.

Now, Bayesians have thought about this problem. One response is the following. You should only consider the finest available partition of the possibility space. So, if there are 100 exclusive diseases (and you know you have one of them), then according to the Principle of Indifference, the probability you have any particular one is $\frac{1}{100}$.

But, first, why? Evidential parity is evidential parity. If evidential parity implies equiprobability, it ought to regardless of the chosen partition.[2] Second, suppose we grant that we should only consider the finest available partition of the possibility space. Returning to our disease with mystery base rate, that rate might take any real value between zero and one. Hence the finest partition is infinitely fine; it contains a cell for every possible rate r. Presumably, the Principle of Indifference recommends a uniform density over that partition.[3] But this doesn't help. It makes things worse; the problem transforms into Bertrand's paradox.

According to the uniform density, the probability is 1/2 that the base rate is less than 1/2. But a uniform density over the base rate entails a nonuniform density over the *squared* base rate. It entails that the probability is greater than 1/2 that the squared rate is less than 1/2.[4] But why should this be? What justifies confidence, in ignorance, that the squared rate is less than 1/2? More importantly, why assign a uniform density over the rate, but not over the squared rate? After all, we have exactly as much evidence that

[2]Many people object at this point. Surely Laplace did not intend that *wherever* you have evidential parity you have equiprobability. Fair enough. You can see this as an invitation to specify exactly what the Principle of Indifference *is* such that it avoids this consequence.

[3]According to the uniform density it is equally probable that r is in the interval (a, b) and in (c, d) just when (a, b) and (c, d) are of the same length.

[4]$P(r^2 \leq \frac{1}{2}) = P(r \leq \frac{1}{\sqrt{2}}) > P(r \leq \frac{1}{2})$ because $\frac{1}{\sqrt{2}} > \frac{1}{2}$.

the squared rate is less than 1/2 as we do that the rate is. A similar problem arises for the cubed rate, the quadrupled rate, the square root of the rate, etc.[5]

Now, Harold Jeffreys and Edwin Jaynes did offer (what they took to be) a solution to this problem. But that solution requires adopting an unbounded or improper distribution as your prior. The sum (technically, integral) of the probabilities of each rate hypothesis is infinite.[6] But an unbounded distribution is not a probability distribution. Probabilities are bounded; they are always less than one.

Other Bayesians have taken another tack. On their view, Bertrand's paradox reveals only that there is no objective starting point in inductive inference. Instead, they contend, "the prior distribution from which a Bayesian analysis proceeds reflects a person's beliefs before the experimental results are

[5]Bas van Fraassen made this problem famous in the philosophical community with his cube factory example (van Fraassen, 1989, 302-307). The problem was originally noticed by Bertrand (1889).

[6]Jaynes recommends (Jaynes, 1968, 20) and Jeffreys considers (Jeffreys, 1961, 123-125)

$$f(p) \propto \frac{1}{p(1-p)}$$

as the prior representative of "total confusion or complete ignorance" (Jaynes, 1968, 20) when an unknown parameter p is restricted to lie between 0 and 1. But the integral of that function does not converge. (Because the integral of $\frac{1}{p}$ diverges over (0, 1), and because $0 \leq \frac{1}{p} \leq \frac{1}{p(1-p)}$ over (0, 1), the integral of $\frac{1}{p(1-p)}$ must also diverge over that interval.)

Jeffreys attributes the view that $f(p)$ is appropriate to J.B.S. Haldane (Jeffreys, 1961, 123), but notes that, for example, were the first two people we investigated to have a disease with unknown base rate, our posterior credence that the population wide rate is 1 would be 1. In other words, we would be certain that everyone has the disease. Jeffreys writes, "The rule $\frac{1}{p(1-p)}$ ... would lead to the conclusion that if a sample is of one type with respect to some property, there is probability 1 that the whole population is of that type"(Jeffreys, 1961, 124). After considering some alternatives, and also finding them wanting, Jeffreys ultimately concludes "we may as well use the uniform distribution ... in the present state of knowledge, that is enough to be going on with"(Jeffreys, 1961, 125). This is just to say that Jeffreys ultimately does not offer a solution to the version of Bertrand's paradox I've presented here.

known. Those beliefs are subjective, in the sense that they are shaped in part by elusive, idiosyncratic influences, so they are likely to vary from person to person." And "trying to force this ... entirely legitimate diversity of opinions into a single uniform one is misguided Procrusteanism."(Howson & Urbach, 2006, 237)

But I do not think this helps with Bertrand's paradox. The problem is not an interpersonal problem. The problem is not that some people want to place a uniform density on the rate, while others want to place a uniform density on the squared rate. The problem is intra-personal. When I ask myself, 'should I adopt a uniform density on the rate, or on the squared rate?' the "elusive, idiosyncratic influences" on my beliefs which are supposed to choose between them do not yield a verdict. Both seem to equally well reflect my evidential station. But I cannot adopt them both; they are inconsistent.

In sum, the question of how to represent ignorance in the Bayesian framework is a vexed one. And I know of no satisfying answer to it. These are not new points; this is just the problem of the priors, the problem that made the Reverend Bayes himself reluctant to be a Bayesian.[7]

[7]Fisher, at least, thought so

> Bayes' introduction of an expression representing probability *a priori* contained an arbitrary element, and it was doubtless some consciousness of this that led to his hesitation in putting his work forward. (Fisher, 1956, 17)

2.2.2 Likelihoodism

I turn then to likelihoodism. Likelihoodists[8] abandon the central Bayesian dogma. Some evidential states – for example, empirical ignorance – cannot be captured by a probability distribution. Sometimes no prior is available.

But to give up priors comes at a cost. Bayesianism allows us to model the acquisition of evidence via Bayes' Theorem. We can condition on our evidence and arrive at posterior probabilities. Thus, we can say how probable it is you have a disease, if you test positive for it in ignorance of the base rate. But the machinery of Bayes' Theorem runs on a prior; it cannot get started if you do not supply one.

Likelihoodists accept this. And they conclude that because we sometimes lack priors, we cannot in those cases locate posterior probabilities. Here is Edwards, accepting that consequence with relish

> It is indeed true that [likelihoodism] ... does not make any assertion about the probability of a hypothesis being correct [in light of the data]. And for good reason: the [view] has been developed by people who explicitly deny that any such statement is generally meaningful in the context of a statistical hypothesis. (Edwards, 1972, 33)

So consider again our disease with mystery base rate. The likelihoodist accepts the upshot of Bertrand's paradox. To settle on a uniform density on the base rate, instead of on the squared base rate, would be arbitrary. But without a prior, we cannot arrive at posterior probabilities. And hence we

[8] like Hacking (Hacking (1965)), Edwards (Edwards (1972)), and Sober (Sober (2008)) (in certain moods).

cannot answer the question of how probable it is you have the disease, upon testing positive. The best we can do is consider the relative likelihoods of this or that hypothesis – i.e., the relative probabilities of our data given this or that hypothesis. So the hypothesis that you are afflicted is best, if you test positive, because given you are afflicted, the probability of a positive test is higher than given you are not. But to say that hypothesis is best is not to say you should be confident it's true, nor is it to say that it's probably true.

My main complaint about likelihoodism is not that it's wrong but that it's too austere. Of course we can rank hypotheses by the probabilities each assigns to our observed data. But I don't think that's the question we were interested in.

Imagine, for example, you test yourself for the disease with mystery base rate one million times. Assume, again, true positive and negative rates of 95 percent. Now suppose that nine hundred and fifty thousand of those one million tests are positive. Rationality requires, it seems to me, that you be confident that you have the disease. It requires that you think it highly probable that you have the disease. But likelihoodism does not deliver this result. The likelihoodist will agree that the hypothesis that you are afflicted is best. (After all, that hypothesis assigns the highest probability to your data.) But without a prior, likelihoodism "does not make any assertion about the probability" that you are afflicted. Thus even after nine hundred and fifty thousand positive tests, it is silent on the question of whether you should be confident you have the disease.

Of course, Likelihoodism could be amended. We could insist on certain thresholds – after 100 positive tests, be confident you are afflicted. But those thresholds and their justification need explication. Until that's supplied, the

view looks too thin.[9]

2.2.3 Classicism

I turn now to the classicists, typically represented by the trio of Ronald Fisher, Jerzy Neyman and Egon Pearson.

In philosophical circles, classical statisticians get a bad rap. Howson and Urbach, for example, claim that classical significance tests yield conclusions which "often flatly contradict those which an impartial scientist or ordinary observer would draw."(Howson & Urbach, 2006, 154).

The classicists did make mistakes,[10] but they are often read uncharitably. They were aware of the problem of the priors for Bayesianism.[11] And they were aware that we could interpret evidence via likelihoods.[12] Given that,

---

[9]This is, in essence, the critique of likelihoodism, pressed convincingly and thoroughly, by Greg Gandenberger. Gandenberger (2016)

[10]Interested readers are referred to Sober, Chapter 1, and Howson and Urbach, Chapter 5.

[11]About Bayesianism, Fisher wrote

> Certainly cases can be found, or constructed, in which valid probabilities *a priori* exist, and can be deduced from the data. More frequently, however, and especially when the probabilities of contrasted scientific theories are in question, a candid examination of the data at the disposal of the scientist shows that nothing of the kind can be claimed. (Fisher, 1956, 17)

[12]In certain places, Fisher seemed to explicitly endorse Likelihoodism:

> Although some uncertain inferences can be rigorously expressed in terms of mathematical probability, it does not follow that mathematical probability is an adequate concept for the rigorous expression of uncertain inferences of every kind. ... More generally, a mathematical quantity of a different kind, which I have termed mathematical likelihood, appears to take [the place of probability] as a measure of rational belief. (Fisher 1935, 474, quoted in Lehmann 1993)

I think we should read them as trying to push likelihoodism further than its austere beginnings. They wanted to see how far they could get without having to invoke prior probabilities. Now perhaps they didn't get very far, but at least some of the criticisms leveled at classical methods fall flat when those methods are viewed in that light.

So what is classicism? I want to abstract away from the details as much as possible. But consider again our disease with mystery base rate, and our test, whose true positive and negative rates are both 95%:

$$P(+|D) = .95$$
$$P(-|\neg D) = .95$$

Here is a general fact that the classicists noticed. Let the unknown base rate be r, and ask: supposing you are about to take the test, how probable is it that you will get an *accurate* result? It's surprising, but that question has a fixed and knowable answer, even though the base rate is unknown. For the test is accurate just in case you test positive, and have the disease, or test negative and lack it. And the chance of *that* is given by

$$\begin{aligned} P((+ \wedge D) \vee (- \wedge \neg D)) &= P((+ \wedge D)) + P(- \wedge \neg D) \\ &= P(+|D)P(D) \\ &\quad + P(-|\neg D)P(\neg D) \\ &= .95 \times r + .95 \times (1 - r) \\ &= .95 \end{aligned}$$

Thus, you do not know how probable it is that a positive test will be accurate. And you do not know how probable it is that a negative test will be accurate.

But you do know, *in general*, how probable it is that you will receive an accurate result. Roughly put, if one billion people took the test, and everyone trusted her result, 95 percent of them would be right. And this is true, again, *no matter what the base rate is*. The point is sometimes put this way: we can know the *pre-data* (or *pre-result*) probability that your test will be accurate, even though we do not know the *post-data* (or *post-result*) probability that it was.

Now, at this point, a rift opens up between our principals. Fisher stands on one side of it, and Neyman and Pearson stand on the other. Allow me to describe their views and my misgivings about each in turn.

2.2.3.1 Fisher

I begin with Fisher. As I read Fisher, he wanted to simply pivot on the accuracy of the test and adopt it as a posterior credence. If, in other words, you test positive, Fisher thought you should be 95 percent confident that you have the disease. And he seemed to argue for this as follows: we have no idea what the base rate is. But we do know that the test will be accurate 95 percent of the time. And in ignorance, it seems rational to *fall back* on the fact that 95 percent of tests are accurate – to use that to determine our posterior credences.

Consider, Fisher might offer, an analogy. Suppose you select Bob at random from a group of people, 95 percent of whom have a heart condition. You are thus 95 percent confident that Bob has a heart condition. But you then learn that Bob is from San Diego. In light of that information, how confident should you be that Bob has a heart condition? Should you, in other words, revise your credences?

It's less obvious, but a base rate problem arises here. To know the probability that Bob has a heart condition, it seems you need to know the rate of that heart condition *among San Diegans*. Or at least, you need to know the rate among the San Diegans in the group from which Bob was selected. If they all have it, the probability Bob does is one. If half have it, the probability Bob does is one half. That said, in ignorance of the proportion of San Diegans who are afflicted, it seems reasonable to fall back. It seems reasonable to let the fact that Bob was selected from a group of people – 95 percent of whom are afflicted – guide and determine your credences.

Fisher wanted to say the same thing about accuracy and our test results. We *know* that 95 percent of results are accurate. It's true that we do not know what proportion of the *positive* tests are accurate. But, in ignorance, it seems perfectly reasonable to fall back, and let the 95 percent accuracy of the test guide and determine our credences. Hence, if you test positive in ignorance of the base rate, you ought be 95% confident you have the disease. [13]

Fisher's view is under-appreciated. He is offering us a middle ground

[13]This is, I think, the easiest way to understand Fisher's *fiducial* argument. He noticed that, in certain circumstances, we could make general probability statements which are true regardless of prior probability distributions. He considers, for example, a random variable X which follows an exponential distribution with rate parameter θ, and notes that the quantity $2\theta X$ will follow a χ^2 distribution regardless of the value of θ. From that, it follows that, if $\chi^2(P)$ is the P-th percentile of a χ^2 distribution, θ will exceed $\frac{\chi^2}{X}$ with probability P.

He then writes,

> The probability statement [– that θ will exceed $\frac{\chi^2}{X}$ with probability P –] had as a reference set all the values of X which might have occurred in unselected samples for a particular value of θ. It has, however, been proved for all values of θ, and so is applicable to the enlarged reference set of all pairs of values (X, θ) obtained from all values of θ. **The particular pairs of values θ and X appropriate to a particular experimenter certainly belongs to this enlarged set,** and within this set the proportion of cases satisfying

in the debate between likelihoodists and Bayesians. We can agree with the likelihoodist that the problem of the priors is fatal to Bayesianism. But at the same time, we can deny the likelihoodists' pessimistic conclusion – that posterior probabilities are out of reach. In fact, we *can* locate and endorse posterior probabilities. We need only find tests which are accurate with a known probability. And our posterior credences can then be guided by the accuracy of those tests.

Now, some authors have claimed that this doesn't really count as progress. Fisher is implicitly invoking a prior, they contend, even if he is unwilling to admit it. (Edwards (1972, 208-209) and Bulmer (1967, 179) both

---

the inequality

$$\theta > \frac{\chi^2(P)}{X}$$

... is certainly equal to the chosen probability P. (Emphasis added.) (Fisher, 1956, 54)

The comment I emphasized above is, I think, crucial. When Fisher refers to the pairs of values "appropriate to a particular experimenter", he just means that the general probability statement is true regardless of the experimenter's prior. In other words, no matter what your prior over θ is, before the data X comes in, you should think that θ will exceed $\frac{\chi^2(P)}{X}$ with probability P.

Fisher then goes on to write,

> It might have been true ... that in some recognizable subset of pairs (X, θ) ... the proportion of cases in which θ exceeds $\frac{\chi^2(P)}{X}$ should have had some value other than P. It is the stipulated absence of knowledge *a priori* of the distribution of θ ... that makes the recognition of any such subset impossible, and so guarantees that in [the experimenter's] particular case ... the general probability is applicable. (Fisher, 1956, 55)

Here, Fisher is just saying that if you *knew* the appropriate prior probability distribution to adopt over θ, you might, upon observing your data, have reason to abandon the general probability statements, which you recognized as true before the data came in. But, he contends, our ignorance somehow "guarantees" that the general probability statement must remain applicable, after you see the data.

raise versions of this criticism.) After all, if the posterior probability that you have the disease is .95, after a positive test, then the prior probability must have been 1/2.[14]

But while that's right, it's hard for me to see how this amounts to much of a criticism of Fisher. The whole *problem* with the Bayesian approach is that it is impossible to locate a prior probability. At worst, then, Fisher has supplied us with a method by which we can locate a rational prior. We can say, 'the *reason* this prior is reasonable is that, in using it, our credences will be guided by the objective accuracy of our tests. *That* is why you should adopt a prior of 1/2 in this case; it has nothing to do with evidential parity.' And that, I think, would count as progress.

There is, however, a more difficult problem for Fisher's view. And

---

[14]

$$\begin{aligned} P(D|+) &= .95 \\ &= \frac{P(+|D) \cdot P(D)}{P(+)} \quad \text{(Bayes' Theorem)} \\ &= \frac{.95P(D)}{P(+|D) \cdot P(D) + P(+|\neg D) \cdot P(\neg D)} \\ &= \frac{.95P(D)}{.95 \cdot P(D) + .05 \cdot (1 - P(D))} \\ &= \frac{.95P(D)}{.9P(D) + .05} \end{aligned}$$

Hence

$$\begin{aligned} .95 &= \frac{.95P(D)}{.9P(D) + .05} \\ .9P(D) + .05 &= P(D) \\ .05 &= .1P(D) \\ P(D) &= .5 \end{aligned}$$

here's a trivial way to see it. Suppose you have a fair coin. On one side of it, you write 'you have the disease'. On the other, you write, 'you don't.' Then, by the considerations above, it seems you ought to reason as follows. You know that that coin will give you an accurate report 50 percent of the time, whether you have the disease or not. So if you flip it, and it says you have the disease, on Fisher's view you ought to adopt a posterior credence of 1/2 that you do. (Same goes, I suppose, if it says you don't have the disease.)

At first blush, maybe that doesn't seem so bad. It looks like an application of the Principle of Indifference. But now imagine you also have a fair three-sided die. On one side, you write 'you have the disease', on another, you write 'you don't and it will rain tomorrow', and on the third, you write 'you don't and it won't rain tomorrow.' (Assume you are ignorant both about the climate, and about whether you have the disease.) You roll the die, and it reads 'you have the disease.' Well, again, regardless of what the truth is, you know the die will report that truth with probability 1/3. And so, by the considerations above, it seems you ought to adopt a posterior credence of 1/3 that you are afflicted.

And the problem is that neither the coin nor the die gives you any evidence. So imagine one person flips the coin and reads 'you are afflicted', and another rolls the die and reads the same. Oddly, on Fisher's view, the coin flipper should be more confident that she is afflicted than the die roller.[15]

[15]And we can make this worse. Imagine a fair million-sided die. And suppose you have a random number generator which generates whole numbers between 1 and 999,999. On one side of the die, you write, 'you have the disease,' on another you write, 'you don't, and this random number generator will generate 1', on another you write, 'you don't, and this random number generator will generate 2', etc. Well then *that* die will again be accurate once every million rolls. And so, if the die says 'you have the disease', on Fisher's picture your posterior credence that you do ought to be one in one million. But again, the die

If we *do* allow our credences to be determined by these mechanisms, problems will percolate into our epistemic futures. Suppose I flip the coin, and arrive at a 1/2 credence I have the disease. You roll the die, and arrive at a 1/3 credence you do. Suppose we then both conditionalize on the same evidence – a collection of positive tests, say. Then we will arrive at different conclusions about the probability we are afflicted.

Thus the problem of the priors re-arises for Fisher. We don't actually need to construct these coins and dice, and see how they behave. We already know that they will provide no further evidence than we already have. And so, if we were to let them guide our credences, we would need to arbitrarily choose a starting point.

---

supplies no evidence one way or another. So if you weren't already, you should not become 99.9999% confident that you do not have the disease, if and just because that million sided die reported you do.

[16]Sophisticates might try to save Fisher by saying something like the following: if you read Fisher, you will find that his whole view rested on the notion of a *pivotal quantity.* A *pivotal quantity*, recall, is a function of the possible data and hypotheses which is independent of which hypothesis is true. If, for example, X is a normally distributed random variable with unit variance and with hypothesized but unknown mean μ, then $(X-\mu)$ is a pivotal quantity – the probability that it takes any particular value is independent of the value of μ.

And the sophisticate will rightly point out that I have not said anything about pivotal quantities here. But, in fact I have been discussing pivotal quantities, just not by that name. Consider again our fair coin; on one side of it, it says 'you have the disease', on the other, it says 'you do not.' And now consider the proposition *the coin reports the truth.* That proposition is a function of the possible data – what the coin reports – and the hypotheses – whether or not you have the disease, into the set $\{\textsc{True}, \textsc{False}\}$. And the probability that that function takes the value True is 1/2, regardless of whether or not you have the disease. Hence that proposition *is* a pivotal quantity, and hence on Fisher's view you ought to pivot on it, and use it to arrive at posterior probabilities.

2.2.3.2 Neyman and Pearson

I turn, lastly, to Neyman and Pearson. Like Fisher, Neyman and Pearson were impressed that we can know the pre-result probability that a test will be accurate, though we do not know the post-result probability that it was. But they were less committal than Fisher about the upshot of that fact.

If Fisher is like a Bayesian, then Neyman and Pearson are like the likelihoodists. They thought that in ignorance of the prior, we should give up on posterior probabilities. Instead, they contended, we should focus on "rules of behavior" that we will follow upon observing our data. "Without hoping to know whether each separate hypothesis is true or false," they wrote, "we may search for rules to govern our behavior with regard to them, in following which we insure that, in the long run of experience, we shall not too often be wrong."(Neyman & Pearson, 1933, 291)

One rule of behavior, for example, is the following. If you confront a positive test, accept that you have the disease. Otherwise accept that you do not. (Let us, along with Neyman and Pearson, leave the notion of 'acceptance' vague.) If you follow that rule, then when you take our test for the disease with mystery base rate, there is a 95% chance that you will accept the truth on any given occasion.

But for Neyman and Pearson, the probability of a hypothesis is always relativized to a rule of behavior. You cannot say categorically, as Fisher or the Bayesians wanted to, 'the probability I have this disease is 95 percent.' For there are often many rules available to you. And there is no fact of the matter about which rule you are following on a particular occasion. Here, for example, are two rules you might follow upon confronting a positive test:

1. Accept that you have the disease only if you test positive, otherwise, accept that you do not have the disease.

2. Accept that you have the disease regardless of your test result.

Notice that to follow each of these rules is to accept that you have the disease, given you've tested positive. But while following the first will lead you to accept the truth 95 percent of the time, following the second will not. So the question is: is accepting that you have the disease *on this particular occasion* to follow the first or the second rule? Depending on how we answer that question, we will be led to different conclusions about the probability that you have accepted the truth. (This is just the reference class problem.)

As with likelihoodism, I do not think Neyman and Pearson's broad view is wrong. The commitments of the view are *right*. It is true that if we follow these rules, we will in fact accept the truth with some known probability (relative to repeated followings of that rule).

But imagine again you test positive nine hundred and fifty thousand times in one million tests. Then it's true that there is a rule of behavior, which you could follow, and which will very, very often lead you to accept the truth. That rule is this one:

(1) If you test positive 950,000 times in 1 million tests, accept that you have the disease, and if you test negative 950,000 times, accept that you do not.

But so what? Should we, in light of this rule, then be confident that you have the disease or not, and on what grounds? Or is there nothing definitive to

be said here? Can, perhaps, the question of how confident we should be *itself* only be answered relative to a rule of behavior?

The intuitive thing to say is this. If you are following a rule which will lead you to accept the truth $p\%$ of the time, then you ought to be $p\%$ confident that you have accepted the truth on this occasion. But that is Fisher's view. Fisher, recall, wanted to pivot on the accuracy of a test to arrive at posterior probabilities. And as we saw, Fisher's view yields inconsistent prescriptions.[17]

Now, as with likelihoodism, Neyman and Pearson's view could be amended. We could insist that, when you follow a rule which will lead you to accept the truth at least p percent of the time, be confident that you've accepted the truth on this particular occasion. But also, as with likelihoodism, that amendment needs explication and then justification.

In sum, in spite of valiant effort, the classicists didn't make much progress on the original problem. Fisher inherits the Bayesian problem of the priors. And Neyman and Pearson inherit the likelihoodist problem of austerity.

2.3 Construction

2.3.1 Introduction, Redux

I turn finally to the constructive part of this paper, to my answer to the question: how confident should you be that you have a disease, if you test

[17]The main problem I raised for Fisher is also awkward for Neyman and Pearson. If you flip a fair coin, one rule of behavior available to you is this one: accept that you have the disease if and only if the coin yields heads. And following that rule will lead you to accept the truth fifty percent of the time. But it is unclear how or whether that fact should bear on your confidence that you are afflicted.

positive for it in ignorance of the base rate?

I want to begin by noting that, in the face of one positive test, it seems reasonable to reassure oneself thus: 'It's true I tested positive. But it's also possible this disease is rare, in which case I needn't worry.'

There are two lessons I want to draw from this.

First, when we are thus reassured, we are focused on a prior probability. To say the disease is rare is just to say that the prior probability you have it is low. But our focus is an objective prior. It must be. You cannot reassure yourself with a subjective prior: 'It's true I tested positive. But perhaps I was antecedently very confident that I didn't have this disease.' That thought is not coherent. Thus I think the question we want answered, when considering this problem is, 'what is the objective probability I am afflicted, in light of my data?' That is the question that I think we care about.

But notice that the objective probability I am afflicted, in light of my data, is inescapably sensitive to the objective prior probability that I was afflicted. In our case, the probability I have a disease is inescapably sensitive to the base rate. And hence – if we're going to answer the question I think we want answered – we must somehow marshal our data in the service of determining objective priors.[18] We must, in other words, somehow marshal our data in the service of determining the base rate. That is the first lesson.

Second, I take it for granted that, at a certain point, it becomes unreasonable to be reassured by the thought that the base rate could be low. Supposing you test positive 950,000 times in 1 million tests, it would be odd

[18]The discussion to follow puts things in terms of base rates, and objective chances. But I think the argument can apply to any notion of objective probabilities you favor – for example, the evidential probabilities of Williamson (2000, 209-230).

to reassure yourself, 'perhaps the base rate is *minuscule* or even *zero*. And thus I needn't be worried.' It is true, of course, that if the base rate is zero, you shouldn't be worried. The posterior chance you are afflicted is also zero. But I take it as a datum that in ignorance that response is too optimistic. And I think our view must account for this transition. It must explain why it might be reasonable to be reassured after one positive test, that the base rate could be low, but not after 950,000.

2.3.2 The Proposal

So here is my proposal. I want to begin by first considering the easier case: suppose you test positive 950,000 times in 1 million tests. This thought, I claim, is too optimistic: 'Perhaps the base rate is *minuscule* or even *zero*. And thus I needn't be worried.' But why is that thought too optimistic?

First, to make things simpler, let us suppose the base rate is zero. Then I contend there is something objectionably *remarkable* about 950,000 positive tests. The base rate's being zero would entail that your epistemic life is more exciting than it probably is.

Here is what I mean. If the base rate is zero, and you test positive 950,000 times in 1 million tests, then you've received a mountain of evidence in support of a hypothesis which had no chance of being true. But now consider the following question: before you observed your data, how confident should you have been that evidence as strong as yours *would* support a hypothesis which had zero chance of being true?

The answer is: you should have thought that extremely improbable. I'll get in to the details below, but for now I hope it's intuitive that 1 million tests, like the one we've described, are very unlikely to provide evidence as strong as

950,000 positive tests in support of a hypothesis that had no chance of being true. For in order to do so, the test would either have to yield 950,000 positive tests, while the base rate is zero, or yield 950,000 negative tests, while the base rate is 1. And the chance of either of those things happening is minuscule.

Now you might object at this point: even if the base rate is *one*, it is unlikely that I would test positive *exactly* 950,000 times (as opposed to, say, 949,487 times, or 950,132 times, or whatever). That's right, but it is a red herring. When I say 'it is unlikely that evidence as strong as that supplied by 950,000 positive tests would support a hypothesis that had zero chance of being true,' I mean that even *conditional* on your receiving evidence as strong as 950,000 positive tests – that is, conditional on your either receiving exactly 950,000 positive or 950,000 negative tests – it is very unlikely that your evidence would support a hypothesis that had zero chance of being true. (Imagine an oracle tells you, 'you will either receive 950,000 positive tests or 950,000 negative tests.' And now ask yourself: how likely is it that my evidence will support a hypothesis which had zero prior chance of being true?)

So that is what I mean when I say that it would be *remarkable* if evidence as strong as that supplied by 950,000 positive tests supported a hypothesis that had zero prior chance of being true.

Now, I need one further claim to get my proposal off the ground. In ignorance of the base rate, I contend, we should not think ourselves or our data remarkable. We should not take, for example, 950,000 positive tests to have supplied evidence that that very data supports a hypothesis which had zero chance of being true in the first place.

I take that last claim to be fundamental. It's meant to be akin to the Principal Principle or the Principle of Indifference. Bedrock. But allow me at

least a bit of rhetoric. When we come into the world, and we confront it in empirical ignorance, nature nevertheless supplies us with an *a priori* guarantee – that our evidence and our epistemic lives will likely be boring and staid. In situations like the one we've been considering, we can be confident that our evidence will support a hypothesis that had some positive probability of being true to begin with. And it would be odd, in ignorance and after the data comes in, to suddenly think you or your data exceptional.

But if that's right, and you test positive 950,000 times, you should be very confident that your data does *not* support a hypothesis that had zero chance of being true to begin with. And hence, if you receive very strong evidence in support of your being afflicted, you should be confident that the base rate is not zero.

That's really the meat of my proposal, though work needs to be done to generalize it. But first let me note that the proposal I'm offering is a classical one, closest in spirit to Fisher's. I suggest we use the *pre-data* fact – that it is very unlikely that we would receive strong evidence in support of a hypothesis which antecedently had zero chance of being true – to be post-data confident, after 950,000 positive tests, that the base rate is non-zero.

But where do we go from here? Even if you've agreed with me thus far, all I've offered is an argument that you should be confident, after 950,000 positive tests, that the base rate is non-zero. But the base rate's being non-zero is consistent with the base rate's being *minuscule*. In fact, it's consistent with the base rate's being so small that, even considering your 950,000 positive tests, it is still objectively unlikely that you are afflicted.

But the generalization step is no giant leap. I've noted so far that it's unlikely that evidence as strong as that supplied by 950,000 positive tests

would support a hypothesis which had zero chance of being true. We can generalize the point by noticing that, as far as our test is concerned, it's also unlikely that evidence as strong as that supplied by 950,000 positive tests would support a hypothesis which had a *minuscule* or very small chance of being true. After all, in order for that to occur, you would either (i) have to test positive 950,000 times while the base rate is very small or (ii) test negative 950,000 times while the base rate is very large. But if the base rate is very small, then it is very unlikely that you have the disease. And hence it is very unlikely that you would test positive 950,000 times. Similarly, if the base rate is extremely high, it is very likely that you are afflicted. And hence it is very unlikely that you would test negative 950,000 times. And thus *in general*, in a situation like this one, it is very unlikely that evidence as strong as that supplied by 950,000 positive tests *would* support a hypothesis that had a minuscule chance of being true in the first place.

In slogan form, then: our evidence will likely support a hypothesis which was likely to be true to begin with. And again, that is a *pre-data* claim. But I suggest we use it to regulate our post-data confidence – to be confident that the evidence we in fact observe also supports a hypothesis which was likely to be true to begin with.

Now, perhaps you're wondering: how is this an improvement on Fisher's view? Here's one way: Recall the fair coin we used to bring Fisher down – on one side of it, it says, 'you are afflicted', on the other, it says, 'you aren't'. Fisher's view entailed that, because the coin's report is accurate 50 percent of the time, we ought to be 50 percent confident that what it reports is true. That's not my view. On my view, we should ask: if I flip this coin, what is the probability that I will receive evidence in support of a hypothesis which

had a very low chance of being true? And the answer is: the probability of that is zero, because the coin is guaranteed not to provide any evidence at all. And so, on my view, after flipping the coin, you ought to be certain that your data does not supply evidence in support of a hypothesis which had a very low chance of being true. But of course you *should*, in fact, be certain of that, because the coin doesn't supply evidence in support of any hypothesis at all.

The second crucial difference, between my view and Fisher's, is that my view targets prior chances. Fisher thought Bayesian subjective priors incoherent, and so tried to offer a view which relied only on probabilistic claims that would be true irrespective of priors. So, in ignorance of the base rate, recall, he thought we should fall back to claims about the accuracy of a test. But likelihoodists, at least, would just raise the following simple objection to Fisher: a test might be accurate, say, 95 percent of the time, and yet indicate a hypothesis is true which is very unlikely to be true. If, for example, the base rate is *zero*, and you test positive using a test which is accurate 95 percent of the time, then it is yet very unlikely that you have the disease. On my view, however, the whole point is to try to use our data to say something *about the prior chances.* Because ultimately the question I think we're interested in is the objective chance we are afflicted, conditional on that positive test. So, if in response to my view the likelihoodist points out that the base rate could be zero, and yet I could test positive, I will respond 'you're right, that's the point! In any given case, there's as much as a 5 percent chance that such a thing could occur – that is, that I should observe data which supports a hypothesis which has zero chance of being true. Thus while I'm fairly confident that the base rate is non-zero, in light of my positive test, I can only be 95 percent confident that that is so.'

2.3.3 Some More Detail

Let me, for thoroughness, work through the details of how the view I've offered responds to our original example. I'll then turn to the question of how to generalize the picture.

Suppose you test positive once in ignorance of the base rate, and your test has fixed true positive and negative rates of 95 percent:

$$P(+|D) = .95$$
$$P(-|\neg D) = .95$$

Now, consider the possibility that the base rate might be, say, less than 1 in 1000. If so, then your evidence supports a hypothesis which had a fairly low chance of being true.

So here is the question I think is relevant: prior to taking the test, what was the probability that evidence as strong as yours should have supported a hypothesis which had less than a 1 in 1,000 chance of being true?

In order for your evidence to do so you would need to test positive while the base rate $r < .001$ or test negative while the base rate $r > .999$.[19] And the probability of *that* is given by

$$\begin{aligned} P((+ \wedge r < .001) \vee (- \wedge r > .999)) &= P(+ \wedge r < .001) + P(- \wedge r > .999) \\ &= P(+|r < .001)P(r < .001) \\ &\quad + P(-|r > .999)P(r > .999) \end{aligned}$$

[19] I take it for granted here that, whatever our account of evidential strength, a positive test is as strong evidence that you have the disease as a negative test is that you lack it.

Considering the left term, $P(+|r < .001)$, note

$$\begin{aligned}
P(+|r < .001) &= P((+ \wedge D) \vee (+ \wedge \neg D)|r < .001) \\
&= P(+ \wedge D|r < .001) + P(+ \wedge \neg D|r < .001) \\
&= P(+|D \wedge r < .001)P(D|r < .001) \\
&\qquad + P(+|\neg D \wedge r < .001)P(\neg D|r < .001) \\
&= P(+|D)P(D|r < .001) + P(+|\neg D)P(\neg D|r < .001) \\
&= .95P(D|r < .001) + .05P(\neg D|r < .001) \\
&= .95P(D|r < .001) + .05(1 - P(D|r < .001)) \\
&= .9P(D|r < .001) + .05 \\
&< .9 \times .001 + .05 \\
&= .0509
\end{aligned}$$

By a symmetric argument

$$P(-|r > .999) < .0509$$

Hence

$$\begin{aligned}
P((+ \wedge r < .001) \vee (- \wedge r > .999)) &< .0509P(r < .001) + .0509P(r > .999) \\
&= .0509(P(r < .001) + P(r > .999)) \\
&< .0509
\end{aligned}$$

Thus, you ought to think, prior to taking the test, there is at most a 5.09% chance that your data will support a hypothesis which has less than a 1 in 1,000 chance of being true. And hence I think it would be mildly remarkable if it did.

And on my view, you ought not take a positive test to be evidence that your data is remarkable. And thus, if you test positive in ignorance of the base rate, you ought be 94.91% confident that the base rate is at least 1 in 1000. And hence, you ought be 94.91% confident that the chance you are afflicted, conditional on your positive test, is at least 1.7%.[20]

Now that is a weak thing to say. But importantly it's not as weak as what the likelihoodists and Neyman and Pearson say. Because as more data comes in, we'll be able to say more. Suppose, for example, you test positive 10 times. Then the chance of receiving evidence as strong as that in support of a hypothesis which had, say, less than a 1 in 1 billion chance of being true is itself approximately 1 in 1 billion. And hence on my view, if you test positive 10 times, you ought to be very confident that the base rate is at least 1 in 1 billion. And hence you ought to be very confident that the posterior chance you are afflicted is high. The view thus captures the natural thought, that after a single positive test, you should be uncertain about the chance you are afflicted, but that after 10 positive tests, or 950,000, you ought to start worrying.

[20]This just follows from Bayes' Theorem, and the fact that it is a monotone increasing function of the prior. If $c(D) > .001$, then

$$\begin{aligned} P(D|+) &= \frac{P(+|D)P(D)}{P(+)} \\ &= \frac{.95P(D)}{P(+|D)P(D) + P(+|\neg D)P(\neg D)} \\ &= \frac{.95P(D)}{.9P(D) + .05} \\ &> \frac{.95 \times .001}{.9 \times .001 + .05} \\ &= 1.7\% \end{aligned}$$

2.3.4 Two further examples and a gesture at generalization

Before closing this essay, I want to discuss two further examples, to stave of some confusion, and to see how the view I've offered here might generalize more widely.

Let us, finally, leave our disease with mystery base rate behind. Instead, imagine the following: you are standing in a circle, surrounded by 1000 cups, one of which contains a ball. You do not know which. And suppose that before you is something that looks like a compass needle. And that needle behaves as follows, if you spin it, there is a 95 percent chance it will land on the cup containing the ball, and a 5 percent chance it will indicate some other cup. (It misfires randomly, say.) You spin the needle and it indicates cup 17.

Here we have 1000 hypotheses about the ball's location, and 1000 possible pieces of evidence. So now let us ask, as we did above: prior to observing your evidence, how confident should you be that your evidence will support a hypothesis that was unlikely to be true, say, that had a 1 in 1000 or less chance of being true?

Notice a problem arises here, which did not arise when considering our disease with mystery base rate. Suppose the ball was placed randomly, and hence the prior chance that the ball would end up in any particular cup was 1 in 1000. (In other words, suppose the chance distribution over the space of cups was uniform.) Then, no matter which cup contains the ball, it was *certain* that you would receive evidence as strong as your evidence in support of a hypothesis that antecedently had a 1 in 1000 or less chance of being true.

The moral is that, in this case, *you cannot place an upper bound* on the probability that you should have observed evidence, as strong as yours,

in support of a hypothesis that had a 1 in 1000 or less chance of being true. But this is not really a problem for the view I've offered here, for we can still find upper bounds, the prior chances we consider just have to be smaller. For example, we can ask, what are the chances that I should observe evidence in support of a hypothesis that had less than a 1 in $10,000$ chance of being true? And by a crude calculation,[21] the chance of that is less than 1 in 10. And hence, on my view, if the needle indicates cup 17, you ought to be 90% confident that the prior chance the ball would be placed in cup 17 was greater than 1 in 10,000.

More generally, the point is that the chance – that you will receive evidence supporting a hypothesis which is antecedently unlikely to be true – depends upon the possible evidence you might observe. But that is our lot. When more possible pieces of evidence are available to us, the more possibilities there are for our evidence to support a hypothesis which was unlikely to be

[21] Let $\uparrow i$ be the proposition 'the needle indicates cup i' and let c_i be the chance the ball is in cup i. Then the chance that our evidence will most strongly support a hypothesis that had less than a 1 in 10,000 chance of being true is given by

$$
\begin{aligned}
&P\Big((\uparrow 1 \wedge c_1 < \frac{1}{10,000}) \vee (\uparrow 2 \wedge c_2 < \frac{1}{10,000}) \vee \ldots \vee (\uparrow 1,000 \wedge c_{1,000} < \frac{1}{10,000})\Big) \\
&= 1000 \times c\Big(\uparrow 1 \wedge c_1 < \frac{1}{10,000}\Big) \\
&= 1000 \times c(\uparrow 1 | c_1 < \frac{1}{10,000}) c(c_1 < \frac{1}{10,000}) \\
&= 1000 \times \Big(c(\uparrow 1|1) c(1|c_1 < \frac{1}{10,000}) + c(\uparrow 1|\neg 1) c(\neg 1|c_1 < \frac{1}{10,000})\Big) \\
&< 1000 \times \Big(.95 \times \frac{1}{10,000} + \frac{.05}{999} \times \frac{9,999}{10,000}\Big) \\
&\approx 1000 \times \frac{1}{10,000} \\
&= \frac{1}{10}
\end{aligned}
$$

true at the outset. But that does not mean the strategy I suggested we use above cannot be applied.

Let me consider one last example, this time dealing with continuous hypothesis spaces. Forgive me if the discussion, at this point, is more abstract than it was for the cases above. More detail can be found in Appendix A.

Imagine you have a friend. She is about to throw a dart at a dartboard, and your aim is to determine at which point she is aiming. Now, her aim is fairly accurate, but it's not perfect – and in fact the error in her aim follows a (bivariate) normal distribution, with a standard deviation of 1 inch. (On average, in other words, she misses her target by about one inch.) She throws the dart, and hits the point p. What should you now think about where she was aiming? We now have an infinite set of hypotheses – all the points she could be aiming at – and prior to the experiment, an infinite set of possible data points.

This makes things more complicated. Suppose, for example, you had asked yourself prior to the experiment, what are the chances that I should observe evidence in support of a hypothesis that had zero chance of being true? This is going to depend on how your friend decided on which point she would aim at. Suppose, for example, that she chose which point to aim at via a uniform density over the dartboard. If so, then it would be a *certainty* that you would observe evidence in support of a hypothesis which had zero prior chance of being true. (This, because every hypothesis of the form 'she is aiming at point x' has zero prior chance of being true.) So it's hard to see how the strategy I advertised above is supposed to apply in this case.

So here is my suggestion for a more general strategy in cases like this. I take it that, whatever evidence is supplied by your friend's hitting some point

p, that evidence is captured in the likelihood function, $L(H) = p(p|H)$, over the space of hypotheses, or over the space of possible points she might be aiming at. (Of course, $p(p|H)$ will have to be a density function now, and not a probability function, but this is no great hurdle.)

Now, let us define a *likelihood region*, R_λ thus:

$$R_\lambda = \{H : L(H) > \lambda\}$$

R_λ is, in other words, the set of all hypotheses which assign a probability of at least λ to your friend hitting the point she did, p. In our setup – where your friend is aiming at a point on a dartboard, and her error follows a standard bivariate normal distribution – R_λ will just define a circle, centered at p, and whose radius is determined by λ. (See Appendix A for an illustration.)

And now I think we can follow the strategy I offered above. You can think about drawing a likelihood region as a procedure – one which you might repeat over many iterations of the experiment: Alice selects and aims at some point, throws a dart, and hits the point p_1, you draw a λ-sized likelihood region around p_1. Bob steps up, aims at some possibly new point, hits the point p_2, and you draw a λ-sized likelihood region around p_2. Charlie steps up and hits point p_3, etc. Then here, I think, is the relevant question – what are the chances that, on any given iteration of this procedure, the likelihood region I draw will circumscribe a region which had a very low *prior* chance of containing the point each friend was aiming at?

Obviously, the answer to that question will depend on the size of λ, and on the size of the dartboard. (If you confront a very large dartboard, and you consider a very small likelihood region, then it's possibly quite probable that the likelihood region you will construct will be one which had a low prior

chance of containing the point your friend was aiming at.) But the question will have a determinate answer, which I think we can and should use to regulate our posterior confidences.[22]

2.3.5 Conclusion

Imagine a conversation between Edwards, the likelihoodist, and Fisher, the classicist. Fisher tests positive 950,000 times for a disease with mystery base rate. Edwards laments, 'There's really nothing to be said, Fisher, about the probability you are afflicted. Without a base rate, we don't have a prior, and hence we cannot locate a posterior probability.' Fisher responds, 'But Edwards, I know this many tests will provide an accurate result with an extremely high probability.' Edwards answers, 'That's true, Fisher, but sometimes very accurate tests indicate hypotheses are true which were very unlikely to be true.'

My contribution to this debate comes at the end of that conversation. Edwards is right; sometimes extremely accurate tests *do* support hypotheses that had a small prior chance of being true in the first place. But my point is: *that itself* is unlikely to happen. So while I think we should worry about the possibility, I also think we should only lend it as much credence as it deserves.

In ignorance of the base rate, or, in ignorance of the chance distribution over some space of hypotheses in general, we can nevertheless arrive at confidence that those hypotheses supported by our evidence are true. To do so, all that's required is that we remain confident, after our data comes in, that our epistemic lives are unremarkable ones – that our evidence supports

[22]For more detail on this example, see Appendix A.

hypotheses which were relatively likely to be true to begin with. And as I've noted here, at least before the data comes in, we should be confident that our epistemic lives will be unremarkable in that respect.

Chapter 3

Evidential Decision Theorists Should Two-Box

3.1 Introduction

In this paper, I argue that causal and evidential decision theory recommend the same action in Newcomb cases. I am not the first to defend this view. But the argument I offer here is not Ellery Eells' tickle defense of evidential decision theory, nor is it Richard Jeffrey's ratificationism, nor does it rely on Huw Price's agent probabilities.[1] (It is, however, in the vicinity of all of those arguments.)

3.2 Newcomb Problems

I begin with a Newcomb problem.[2] Imagine you know an expert psychologist. She offers you $1,000, but says, 'earlier today, I read your psychological profile, and predicted whether or not you would take this $1,000. If I predicted you *would*, I did nothing, but if I predicted you *would not*, I deposited one million dollars in your bank account.' Normally you would be skeptical. But you have seen the psychologist accurately predict (and make the appropriate deposits) 1,000 other times with 1,000 other people, with say an equal number of people refusing and accepting the $1,000, and she has only

[1] See Eells (1981), Jeffrey (1983), Price (1991)

[2] This example is J.H. Sobel's by way of Jim Joyce (Joyce, 1999, 146-157).

been wrong once. The question is: should you accept the $1,000?[3]

And the puzzle here is as follows: while your refusing the $1,000 would apparently provide you with *evidence* that you have an additional one million dollars in your bank account, nevertheless refusing is causally inert with respect to the contents of that account. Thus, if you want evidence that you have $1 million in your account, it seems you should refuse the $1,000. But if you are instead guided by the causal implications of your actions – if your aim in acting is to causally bring about positive outcomes – you ought to accept the $1,000. This last, because whether the psychologist deposited $0 or $1 million in your bank account, you would be better off with an additional $1,000.

On the orthodox line, to refuse the $1,000 – in the hopes of receiving evidence that the psychologist deposited $1 million – is to follow the recommendations of evidential decision theory. And to accept the $1,000 – perhaps recognizing that you cannot, in acting, change the past – is to follow the recommendations of causal decision theory.

---

[3]Robert Nozick Nozick (1969) originally presented the problem as follows:

> Suppose a being in whose power to predict your choices you have enormous confidence.... There are two boxes (B1) and (B2). (B1) contains $1000. (B2) contains either $1,000,000 or nothing. You have a choice between two actions
>
> (1) taking what is in both boxes.
>
> (2) taking only what is in the second box.

Furthermore, and you know this, the being knows that you know this, and so on:

(I) If the being predicts you will take what is in both boxes, he does not put $1,000,000 in the second box.

(II) If the being predicts you will take only what is in the second box, he does put the $1,000,000 in the second box.

Formally, *evidential* decision theory recommends that you act so as to maximize, over all available actions A, and for outcomes O

$$EEV(A) = \sum_O P(O|A) \cdot V(O\&A)$$

Here, $P(O|A)$ is just given by the agent's subjective credence function – it is the subjective conditional probability of O given A – and V is the agent's valuation function.

Causal decision theory, on the other hand, recommends that you act so as to maximize, over acts A and for outcomes O:

$$CEV(A) = \sum_O P(A > O) \cdot V(O\&A)$$

Where $P(A > O)$ is referred to as the *causal probability of O given A*. And this is typically cashed out as something like the objective chance of O which would result from the performance of action A. And because the objective chance that you've received \$1 million does not change if you refuse the \$1,000, causal decision theory recommends accepting it.

3.3 Correlation and Causation

The rallying cry of causal decision theorists is that old saw 'correlation is not causation!'[4] The number of people who drown in San Diego is positively correlated with ice-cream sales.[5] Nevertheless, if you are a nefarious ice-cream salesman who cares little about the suffering of others, it is irrational to round

[4]Here's Joyce: "Evidential relevance isn't causal relevance; Correlation isn't causation; indicating is not promoting." (Joyce, 1999, 163)

[5]People both buy more ice cream and go swimming more often in the summer, which is what explains this correlation.

up a bunch of San Diegans and drown them. Or, to consider a Stalnakerian medical example:[6] Suppose you know that IQ is positively correlated with alcohol abuse. (This may be true, actually.) But suppose you also know – as is plausible – that abusing alcohol does not *cause* you to have a high IQ. Then you should not cultivate a drinking habit in the hopes of improving your IQ. Similarly, in the Newcomb problem, refusing $1,000 is correlated with having one million dollars in your bank account. But 'correlation is not causation!' the causal decision theorist cries, and thus it would be foolish to refuse.

That, it seems to me, is right. But what has always struck me as strange about these sorts of examples, levied in support of causal decision theory, is this: cultivating a drinking habit, let's stipulate, will not cause you to have a high IQ. But it's further not at all clear that cultivating a drinking habit – in the hopes of improving your IQ – would provide *evidence* that you have a high IQ either. Imagine, for example, you have a friend who has tried but has never enjoyed alcohol. One day he reads in *Reader's Digest* that alcohol abuse is correlated with IQ, and so forces the habit upon himself. As far as I can tell, you do not, and he does not, thereby get evidence that he has a high IQ. But if that's right, then evidential decision theory, along with causal decision theory, would not recommend his drinking.

You might think the intuition in that case a mere artifact of the chosen example. Here's another one: being an academic is (surely) non-causally correlated with owning a Volvo. Now suppose your friend learns about this correlation, decides he wants to be an academic, and so purchases a Volvo. Again, that purchase, it seems to me, does not provide evidence that that friend is or

[6]See Stalnaker (1980 [1972]), and for discussion Gibbard & Harper (1978 [1981]). An example like this is also discussed in Egan (2007).

will become an academic, despite the correlation between purchasing a Volvo and becoming an academic.

Really these examples are not hard to find: flossing is correlated with heart health. But the connection is not causal; instead, those who take better care of themselves in general will both floss and eat a heart-healthier diet. So if a friend takes up flossing only upon learning of the correlation, and otherwise makes no changes in his life, you do not get evidence that the condition of his heart will improve.

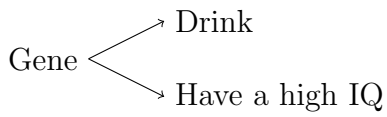
In all of these cases, it seems to me, the problem is that your friend is attempting to exploit a mere correlation to obtain some end – a higher IQ, or an academic career, or a healthier heart. And it just doesn't seem true that, in making that attempt, he receives evidence that he has already or will succeed. I think the same is true in Newcomb's problem. If an evidential decision theorist refuses the $1,000, she is attempting to exploit a mere correlation to obtain evidence that she has $1 million. She is (in a way I will later make precise) like our reluctant drinker who forces the habit on himself in the hopes of improving his IQ. And it's at least not obvious that refusing *would* supply her with the evidence she seeks – that she has $1 million in her bank account. In fact, I think we can *show* that she does not receive that evidence. And if that's right, then evidential decision theory – along with causal decision theory – also recommends her accepting the $1,000.

3.4 Correlations, Knowledge, and the Timeline

In spelling out my argument, I want to begin with the assumption that every genuine correlation arises out of some, perhaps distal, common cause. The reason that drinking is correlated with IQ is, perhaps, because there is

a gene which both disposes people to drink, and equips them with a high IQ. The reason that academics buy Volvos, assuming one is not the cause of the other, is that there is some feature F which causes both Volvos and an academic career to seem appealing to those who have F. In later sections I will consider whether or not that assumption can be relaxed. (Spoiler: I think it can.)

So let us suppose there is a gene, disposing those who have it to drink, and which equips those who have it with a high IQ. Then we can represent the causal structure of the situation in this simple graph:[7]



Now here is, I think, the crucial question: given this causal structure, do people with the gene drink *because* doing so is correlated with having a high IQ?

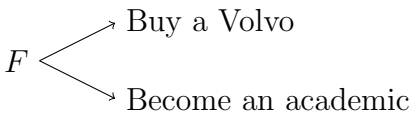
On the most natural understanding of the situation, the answer to that question is *no*. The gene is, after all, *responsible* for the correlation. So the timeline is as follows: the gene causes those who have it to drink and to have a high IQ, thus causing there to exist a correlation between drinking and having a high IQ. We then later come to discover that correlation through observing that drinkers on average have higher IQs than non-drinkers. The important point is that in a natural case like this one, our knowledge of, or belief in, the correlation comes *after* the gene establishes the correlation. From that it follows that those who were driven to drink by the gene were not motivated

[7]Ellery Eells drew pictures that look very much like my pictures Eells (1981). But, as I will discuss below, Eells and I look at these pictures and see different things.

to do so by the belief or knowledge that drinking is correlated with having a high IQ.

So here is a true conditional: if those with the gene did not believe or know that that there existed a correlation between drinking and having a high IQ, they would yet be motivated to drink.

We can say a similar thing about Volvo buying academics. Suppose there is some feature F which disposes those who have it both to Volvos and to the academy. Thus



Then again the most natural timeline is as follows: those with F buy Volvos and become academics. This establishes a correlation, which we later discover. And hence this conditional is true: if those with F did not know or believe that buying Volvos was correlated with becoming an academic, they would yet purchase Volvos.

So let us return to our friend who forces upon himself a drinking habit, only after and, let's say, only because he learns of a correlation between drinking and IQ. Then I think we can explain why his drinking does not provide evidence he has a high IQ as follows: he *would not be inclined to drink* if he did not know about the correlation between drinking and IQ. And thus a conditional which is true of those who have the IQ-alcohol gene, is not true of your friend. Hence your friend does not have the gene, and therefore his drinking is not evidence that he has a high IQ.

The same goes for Volvos and the academy. If there is a feature F which both inclines those who have it to purchase Volvos and to become academics, then – on the most natural understanding of the case – if you have F you would be inclined to buy a Volvo even if you didn't know that buying a Volvo is correlated with becoming an academic. So if you buy a Volvo *only because* doing so is correlated with becoming an academic, you do not have F, and hence your purchasing a Volvo is not evidence that you will become an academic.

The story about Newcomb cases is a bit more complicated, but I'll argue in the next section that the argument applies to them as well. But before that let me make a simpler point: if the existence of a correlation between some action and some outcome were enough to establish that *your* performing that action is evidence that you will secure that outcome, then evidential decision theory was in trouble long before Newcomb's problem arose. For evidential decision theorists would be buying Volvos in the hopes of becoming confident that they will secure academic postings and trying to drink their way to confidence they have high IQs. And that would be objectionable without involving psychologists with preternatural abilities to predict our future behavior.

But, as I've argued here, whatever is the true decision theory, trying to drink your way to confidence you have a high IQ or maneuver a Volvo into an academic career does not provide you with *evidence* that you have a high IQ or will have an academic career. And hence evidential decision theory does not recommend the attempt.

3.5 Back to the Newcomb

I think the point of the above section extends directly to Newcomb problems. But to see how I want to simplify the case.

Forget the expert psychologist for a moment. Instead, imagine that you and 1,000 friends come across a black box – it looks vaguely like an ATM – out in the middle of the desert. And you notice $1,000 protruding from the cash slot. You then watch as each of your friends steps toward the machine and either, after inspecting the machine for a while, leaves the $1,000 in the slot, or takes the $1,000 (after which it is replaced for the next friend). But then, everyone checks their phone, and those who *did not* take the $1,000 notice an email from their bank, thanking them for a recent $1 million deposit. The crazy part is that, in each case, the timestamp on the email reveals it was sent moments *before* the person declined to take the $1,000. So somehow, the box is predicting whether or not you will refuse, and making deposits accordingly. You then step up to the box, and have to decide whether or not to take the $1,000.

I think this counts as a Newcomb problem, but I also think it is directly analogous to the alcohol-IQ, and academic-Volvo cases described above. If there is a genuine correlation here, then there is some feature F of those who refuse the $1,000, which allows the machine to place $1 million into refusers' bank accounts. But notice that those who refused the $1,000 did so *without* knowing or believing that doing so was correlated with having $1 million dollars in their bank accounts.

And really, whenever we come to learn about a correlation through the observation of correlated events, it has to be that way. You can only observe a correlation if that correlation exists. But the correlation exists only

if there is some common cause whose operation preceded the correlated events. [8] Hence that common cause precedes and thus cannot be sensitive to our knowledge of or belief in the existence of that correlation. *Maybe* there are complicated causal structures that we can imagine where the belief that two events are correlated is required, in order for the cause which generates that very correlation to operate. I will consider that possibility later. For now I'll just point out that huge swaths of cases that have been called Newcomb problems do not have a complicated causal structure like that.

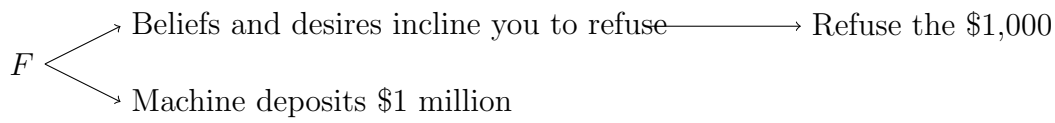
Returning to our ATM, you've watched half of your 1,000 friends refuse, and those who did all noticed they received $1 million deposits. Now, I contend, if you want to know whether or not *your* refusing would provide evidence that you have $1 million in your bank account, the question you face is: if you did not know that refusing was correlated with discovering $1 million in your bank account, would you yet refuse? And if you are an evidential decision theorist, who would behave rationally, the answer to that question is *no*. (After all, in that situation, the expected value of refusing is $0, while the expected value of taking the $1,000 is $1,000.) Hence, if you are a rational evidential decision theorist, you do not have the feature which inclines people to refuse (and which the machine exploits to deposit $1 million) and thus your refusing would not provide you with evidence that the machine has placed $1 million in your bank account.

Actually, I think that last part, though true, is not crucial. Suppose, for example, you do convince yourself that you would refuse the $1,000, even if you did not know about the correlation between refusing and the machine's

[8] Again, I will later try to relax this assumption, because it is contentious.

depositing $1 million. Perhaps you wouldn't take the money, because it would seem to you too good to be true. Then it seems to me *that fact* – the fact that you would refuse – gives you all the evidence you can get that you have the feature F, and hence that the machine deposited $1 million in your bank account. Put more precisely: conditional on the way you would behave if you did not know refusing was correlated with receiving $1 million, refusing is independent of receiving $1 million. And if that's right, then evidential decision theory *still* does not recommend refusing, even if you *would have* refused were you unaware of the correlation. Similarly, if you haven't yet but would have taken up drinking even if you did not know that drinking was correlated with having a high IQ, well then that is evidence you have a high IQ. But then your actually doing so does not provide any *further* evidence of your intelligence.

That last, perhaps, reminds you of Ellery Eells' tickle defense of evidential decision theory Eells (1981). Eells' view, recall, was this: the only way the feature F could cause you to refuse is by first causing you to have certain beliefs and desires which would incline you to refuse. Hence the causal structure is more like this:



But, Eells contended, in a decision problem, you know your beliefs and desires (as represented by your subjective probability and valuation functions). And given the causal structure just sketched, conditional on those beliefs and desires, refusing the $1,000 and gaining $1 million are independent.

But to the extent Eells and I agree, I don't think he captures the whole story here. Because on Eells' view, anyone whose beliefs and desires incline them to refuse will get evidence that they have $1 million in their bank accounts. But now recall your friend who forces on himself a drinking habit only after and because his drinking is correlated with having a high IQ. Given he took up drinking, his beliefs and desires do apparently incline him to drink. But, I contend, he does not receive, nor do those beliefs and desires supply, evidence that he has a high IQ.

Eells is right that – if there is a common cause of drinking and having a high IQ – that cause will operate by influencing a person's beliefs and desires. But the crucial question, I contend, is not whether a person's *current* beliefs and desires incline her to drink, but instead whether she *would be* so inclined, were she unaware that alcohol abuse and IQ are correlated.[9]

3.6 Recursive Newcomb

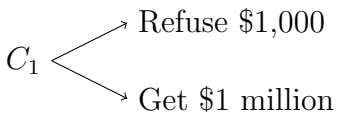
Let us make things a bit more complicated. I promise I will return to our expert psychologist eventually. But first consider again our desert ATM. $1,000 protrudes from its cash slot. But suppose this time you confront it with 2,000 friends, instead of a mere 1,000.

[9]I think that this general point applies to the other reconciliations of causal and evidential decision theory that I mentioned in the introduction – those offered by Jeffrey Jeffrey (1983) and Price Price (1991). Both Jeffrey and Price contend that there is some fact about you, in relation to the decision problem you *currently* face, which renders your refusing independent of your receiving $1 million. For Price, it's the belief that your refusing would provide evidence that you have $1 million in your account. For Jeffrey, it's the fact that, in the decision problem you currently face, you would accept the money. My view is different: I contend that to answer the question of how you ought to behave in the decision problem you *currently* face, you must ask yourself how you would behave in a different decision problem – namely one where you did not believe your action and the desirable outcome correlated.

Half of those 2,000 friends approach the ATM. Some refuse, some do not. The bank sends the relevant emails, and thus you learn about a correlation between refusing the $1,000 and the machine's depositing $1 million. And now you and the 1,000 remaining friends have to decide what to do. You all confer, and conclude that the first wave was mad – there is *no way* you would have refused if you did not know about the correlation. So then, each of these latter 1,000 friends approaches the machine. Then, some of *them* refuse, and some of them do not. And, again, each of the refusers checks her phone and discovers that $1 million has been deposited in her bank account. You then approach the machine, and you have to decide whether or not to refuse the $1,000.

Now, notice, the second wave of refusers would not have refused if they did not know about a correlation between refusing and receiving $1 million. But nevertheless, their refusing *was* correlated with receiving $1 million. So surely *your* now refusing would provide you with evidence you have $1 million in your bank account, even though you would not have refused if you did not know about the correlation.

That seems right, but I don't think it is. The first wave of 1,000 friends establishes the existence of some common cause C_1 of their refusing, and of their receiving $1 million. Thus the causal structure is this one



But the second wave of 1,000 friends, we stipulated, all agreed that they would not be inclined to refuse were they unaware of the correlation between refusing the $1,000 and receiving $1 million. Hence C_1 is not present for any of the

friends in the second wave. Thus, given there is a correlation between second wavers' refusing and their receiving $1 million, there must be some *other* cause $C_2 \neq C_1$, which along with the knowledge of a correlation between refusing $1,000 and getting $1 million, causes them to refuse:

$$[C_2 + \text{knowledge } \exists C_1 \neq C_2] \begin{cases} \nearrow \text{Refuse \$1,000} \\ \searrow \text{Get \$1 million} \end{cases}$$

So now you approach the ATM. As a faithful evidential decision theorist, you know C_1 is not present in your case (because you would not be inclined to refuse if you did not know the relevant correlation existed). But now ask: is C_2 present in your case? Well, if C_2 were present in your case, then you would be inclined to refuse upon learning of a correlation between refusing $1,000 and gaining $1 million. But, recall, we decided above that an evidential decision theorist would not be inclined to refuse upon learning of that correlation! Hence you do not have C_2 either. And hence your now refusing would not be evidence that you will receive $1 million.

We can iterate, and think about a third wave, and a fourth, etc. But, so long as you are committed to always following the recommendations of evidential decision theory, the end result will be a recursive structure. And the question of whether your now refusing is evidence that you will gain $1 million will always ground out in the base case, in your answer to the question: would you be inclined to refuse if you did not know of a correlation between refusing and gaining $1 million? And if you are an evidential decision theorist, the answer will be *no*, and hence evidential decision theory never recommends refusing, no matter how many waves of friends successfully exploit the correlation between refusing and gaining $1 million.

3.7 Objection 1: returning to our expert psychologist

I return, finally, to our expert psychologist. I have been told that all of the above is irrelevant to the Newcomb case that I presented at the beginning of this essay. (Recall its contours: an expert psychologist reads your psychological profile, and predicts whether or not you will refuse her offer of $1,000.) It is irrelevant because – putting things in terms of the recursive structure described above – in that case, *there is no base case.* There is no first wave of participants, which confronts the psychologist without knowing of or believing in her predictive abilities. Instead, every participant believes in the psychologist's predictive abilities, and that belief is perhaps even a condition on her ability to make her predictions. And without such a base case, the recursive structure I described above collapses.[10]

In response let me first reiterate that, even if the charge sticks, the recursive structure I described above will arise whenever we discover a correlation through observation. And that includes any realistic medical Newcomb case, like our alcohol-IQ case. (Another widely discussed one is the smoking lesion case, where a lesion both inclines people to smoke, and gives them lung cancer.) In these sorts of cases the cause which generates the correlation will not require nor be sensitive to the belief that that very correlation exists. I will still call it a victory if all we've here discovered is that evidential and causal decision theory agree in these cases and a wide range of similar cases where people have thought their recommendations diverged.

Second, I would find it unsettling if somehow the *testimony* of the expert psychologist were crucial to generating a true Newcomb problem – that

[10]Thanks to Andrew del Rio, David Sosa, and Miriam Schoenfield for pressing me on this point.

is, to generating a case which reveals an actual distinction in the recommendations of causal and evidential decision theory. Imagine some non-linguistic agents, who can only come to learn about correlations through the observation of correlated events, or who only believe that correlations exist when their evidence suggests they do. The recursive structure just described will always arise for them. I would find it odd if it turned out that we linguistic agents require a distinction between evidential and causal decision theory, while these non-linguistic, but nevertheless wholly rational, agents would not.

Third, when spelling out the Newcomb problem, authors are typically explicit that you *do* receive evidence that convinces you of the psychologist's predictive abilities. Here's Nozick: "you know that the being has often correctly predicted the choices of other people"(Nozick, 1969, 115). And here is Joyce: "You have seen the psychologist carry out the experiment on two hundred people, one hundred of whom took the cash and one hundred of whom did not, and he correctly forecast all but one choice"(Joyce, 1999, 147). I wonder what is the point of introducing these details. But one natural thought is as follows: if *you* had confronted the predictor without observational evidence of her reliability, you would not believe in those predictive abilities, and hence you would not have been inclined to refuse the $1,000. But if that's right, then I will just run the argument above: if you had whatever feature allowed her to predict the initial refusers would refuse – in the absence of any observational evidence of her predictive abilities – you would also be inclined to refuse if you lacked observational evidence of her predictive abilities. But you would not be so inclined, and hence you do not have the feature, and thus your refusing would not supply evidence she will accurately predict your behavior. In other words, in the original presentations of standard Newcomb problems, like

the Newcomb problem rehearsed at the beginning of this essay, the recursive structure does seem to arise, base case and all.

More strongly, it seems to me that it is irrational to believe the testimony of the predictor if you have not independently verified her predictive ability. And if so, then, at least for rational agents, there will have to be a base case – it will consist of those who refuse though they have not verified the psychologist's predictive ability.[11] (What about irrational agents? More on that in the next section.)

One last point: The objector, recall, is proposing that everyone who confronts our expert psychologists believes in her predictive ability, and that that is a crucial difference between that case, and for example, the desert ATM case I presented above. But I think that the objector confronts a dilemma here, which arises from this question: is the *belief* that the psychologist is an accurate predictor required in order for her to *be* an accurate predictor?

Suppose the answer to that question is *no*: whether you believe she can or not, the psychologist can predict whether you will refuse the $1,000. In that case, I will just run the argument as I ran it above. It doesn't change the point if everyone happens to take the predictor at her word. Because you can still ask yourself the question: would you be inclined to refuse if you did not believe the predictor was an accurate predictor? And if you are an evidential decision theorist, the answer will be *no*. But, because her predictive ability is not sensitive to the belief, if you had the feature which both inclines those

[11] This is obviously related to the debate about testimonial *reductionism* – see Adler (2017) for general discussion. If you are sympathetic to reductionism, then you should be sympathetic to the line taken in this paragraph. But you can be an anti-reductionist and agree that, in this special case, it is not rational to accept on faith the predictive ability of the expert psychologist.

who refuse to refuse, and allows the predictor to deposit $1 million, then you would be inclined to refuse. Hence you do not have the feature, and therefore your now refusing is not evidence that she deposited $1 million.

So, in order for the objection to succeed, the answer must be *yes* – the psychologist can predict whether you will refuse the $1,000, but only because you believe she can. We might suppose – to put some meat on the proposal – that our expert psychologist is also an accomplished biologist. And in her research, she has discovered that there is a peculiar gene, which operates as follows. If and only if an agent believes the psychologist can predict whether that agent will refuse the $1,000, *then* if that agent has the gene, he will refuse, and if the agent lacks it, he will not. But without the belief, the gene is inert. In other words, the common cause which both inclines people to refuse the $1,000, and allows the psychologist to deposit $1 million, requires, in order to operate, the *belief* that refusing $1,000 and the psychologist's depositing $1 million are correlated.

But I worry about the cogency of a case like that. It has some very peculiar self-referential features. Suppose, for example, that you are the first person the expert psychologist/biologist informs that she has discovered this peculiar gene, and she offers you $1,000. Then you are in the following situation: you know that if and only if you *believe* that the psychologist can predict your behavior, then she will be able to do so. So you confront a proposition which you can make true by believing it, and make false by disbelieving it. It is as though an extremely trustworthy friend had said to you "I can predict what you will eat for lunch tomorrow, but only if you believe I can." What is the appropriate response if he then asks, "So, do you believe I can predict

what you will have for lunch tomorrow or not?"[12] Or suppose you watch as the predictor approaches 1,000 people, some of them believe that she can predict their behavior, which allows her to do so. But some of them do not believe it, and her accuracy for them is no better than random guessing. She then approaches you and says, "If and only if you believe I can predict your behavior, then I can." She offers you $1,000 and says "If I predicted you would refuse, I deposited $1 million in your bank account." What should you do?

It's hard for me to see how decision theory can even get a foothold in a case like this, because in order for the decision theoretic machinery to operate you need to first pin down your beliefs. You have to first decide whether or not you believe the predictor's testimony. And at least in my own case I would find it very difficult to even answer the question of whether or not I believe the predictor can predict my behavior here. (Have I received evidence that she *can* predict my behavior? That depends on whether or not I believe she can. If I do believe it, then yes I have, but if not, then no I haven't.)

My point is that things start to go awry if a condition on the predictor's predictive ability is your belief in her predictive ability. And on grounds of taxonomic simplicity, I am reluctant to insist that cases like this are coherent and thus reveal some deep distinction in decision theory. I pass the burden to those who want to make our lives more complicated, to show us that such cases require that we do so.

[12]There are related examples in the Epistemic Utility Literature, e.g., Jennifer Carr's Yoga Teacher example (modified for my purposes) Carr (in press): Suppose your Yoga teacher tells you "you can do a handstand, but only if you believe you can." He then asks you "So, do you believe you can do a handstand or not?"

3.8 Objection 2: have we really eliminated the distinction between EDT and CDT?

You might think that everything in this paper misses the point in a somewhat trivial way. For when I introduced the distinction between evidential and causal decision theory, I said that evidential decision theorists were guided by subjective credences, where causal decision theorists are guided by objective probabilities. But subjective credences are radically unconstrained! The starting point of, for example, the dutch book argument, and the accuracy argument, for probabilism is the assumption that we *could* at the same time both be 90 percent confident it's raining, and 90 percent confident it's not. (What the dutch book and accuracy arguments reveal is just that we *should* avoid that situation, at least if we want to keep our money, or if we want our beliefs to be as close to the truth as possible.) And there is nothing inconsistent in imagining a person who both (i) believes her refusing $1,000 will provide evidence she has $1 million in her bank account, and yet (ii) believes that her refusing the $1,000 will not *cause* $1 million to show up in her bank account. So we can just *stipulate* that an agent is so constituted, and it will follow that there will be a distinction between the recommendations of causal and evidential decision theory, for that agent.[13]

That is a fair point. To an extent, I have abandoned subjective credences and thus evidential decision theory as it was originally spelled out by Jeffrey Jeffrey (1983). Recall the alcohol-IQ case. Your friend has never enjoyed alcohol, but forces the habit on himself after learning of a correlation between drinking and IQ. I said he doesn't receive any evidence that he has a high IQ. And I think this would be so even if he does, in drinking, become

[13]Thanks to Sinan Dogramaci for this objection.

more confident he has a high IQ. So the notion of *evidence* I am relying on here is not simply *increase in subjective credence.*

But here, more precisely, is how I would like to be understood. I take myself to have given an argument that it is not *rational* to believe that, e.g., your drinking is evidence you have a high IQ, if the only reason you are drinking is because drinking is correlated with having a high IQ. And similarly, I take myself to have given an argument that it is not rational to believe that refusing the $1,000 is evidence the psychologist deposited $1 million in your bank account.

Of course, someone might have irrational beliefs. And evidential decision theory might recommend that someone with irrational beliefs take an irrational action. But this is not objectionable. If someone irrationally believes that drinking a can of paint will make him happy, then evidential decision theory might recommend that he drink that can of paint. But that does not constitute an objection to evidential decision theory, for it is no surprise that irrational inputs yield irrational outputs. A similar problem, notice, will arise for causal decision theory; if I mistakenly believe that drinking a can of paint is valuable, and there is an action that will cause me to drink that can of paint, then causal decision theory will recommend that I take that action, even though it is irrational to do so.

What the Newcomb problem was supposed to show, I contend, is that two rational agents, apprised of all the same facts, might be led to different conclusions about what they ought to do, depending on whether they are sensitive to the causal or the evidential structure of the world. But if you are rational, I've here argued, the causal and the evidential structure of a Newcomb problem world are the same.

Now, you might keep pressing here. You might say, 'but nevertheless you've admitted that, for some agents – namely those with irrational beliefs – causal and evidential decision theory recommend different actions!' That's right, but I think we can just as well explain that distinction as follows: there is the action that decision theory recommends, given your beliefs, and the action that decision theory would recommend, were your beliefs rational. That captures the distinction, but does not require our introducing the extra machinery of causal decision theory.

3.9 Objection 3: what about QM correlations?

One last objection: recall that everything I said above relied on the principle that every genuine correlation is the result of some common cause. But aren't quantum correlations a counterexample to this principle? Isn't the behavior of entangled particles correlated, even though there is apparently no common cause which underlies that correlation?

The answer to that question depends on your interpretation of Quantum Mechanics.[14] But I do not want the argument I've given above to rely on or require any particular thesis in quantum physics. So let us return to our predictor, and imagine that there is a brute, quantum correlation between her depositing $1 million in people's bank accounts, and their refusing $1,000 when she offers it to them.

Before responding to the problem directly – and I realize I'm becom-

[14]See Ahmed & Caulton (2014) for discussion. They argue that interpretations which do not rely on some common cause underlying a correlation present a problem for causal decision theory. That problem is centered around this question: what does causal decision theory recommend, if there is an objective correlation, but which is not underwritten by any common cause? See Adam Koberinski & Harper (in press) for a response.

ing a bit of a broken record at this point – let me say that I think that even if my argument falls apart here, we will still have discovered something interesting – that, in a Newcomb problem, there is only a distinction between the recommendations of causal and evidential decision theory when there is a brute, uncaused, perhaps quantum mechanical correlation between two events. I think that would be a surprising result. And many cases that people have thought were Newcomb problems are not so constituted – e.g., medical Newcomb problems like the alcohol-IQ case discussed above.

But I think that argument I've given above actually applies equally well here, even if a Newcomb problem is underwritten by a brute uncaused correlation. Though I relied a great deal on the common cause which underlies a correlation, I think that those common causes can be expunged; they can be seen as a way to focus the mind, but they are ultimately unnecessary.

Suppose again you come across an ATM in the desert, and you watch 1,000 people either take or refuse the proffered $1,000. Those that refuse discover that the ATM deposited $1 million in their bank accounts before they refused. But this time you know that the correlation is brute, and not underwritten by any common cause. Nevertheless, if you are an evidential decision theorist, who would follow the recommendations of evidential decision theory, then you know the following fact about yourself: that you would not refuse, if you were unaware of the correlation between refusing and the machine's depositing $1 million. And that supplies you with evidence that the machine did not deposit $1 million in your bank account. And your *now* refusing doesn't change that – you can't exploit the correlation twice, as it were, once to receive evidence that the machine did not deposit $1 million, and a second time to receive evidence it did. Or, relying on our trusty example, suppose there

was a brute, uncaused correlation between alcohol abuse and IQ. Your alcohol hating friend learns about this, and so forces the habit upon himself. The fact that he *would not* have abused alcohol, if were unaware of the correlation, settles the question of whether his now drinking is evidence he has a high IQ. It's not.

The moral is this: when you know what you would do were you unaware of a correlation, that alters the evidential force of your available actions. And while it's easier to recognize this when correlations are understood in terms of common causes, the point stands even in the absence of those common causes.

Appendices

Appendix A

Probably Not that Improbable: More detail on the continuous case

A.1

Here is a more thorough description of how I would like to think about the continuous case discussed in Chapter 2.

Consider again the dartboard example, put in the language of statisticians. Let x be a realization of a random variable X which follows a standard normal distribution with mean μ, where μ is known to lie in a circle C with area A. Here μ corresponds in the original example to the point your friend is aiming at, and x is the point she hits. Let $p(\mu)$ be the objective prior distribution on μ over the points in C. Now, define the likelihood region $R_\lambda(x)$, for each possible realization x of X, as follows:

$R_\lambda(x) = \{\mu \in C \mid \phi(x|\mu) > \lambda\}$ where $\phi(x|\mu)$ is the pdf of a standard normal distribution.

$R_\lambda(x)$, in other words, contains all values of μ which assign a density of at least λ to x. Finally, let b_{R_λ} be a point on the boundary of R_λ. (See Figure 1 below.)

Now, consider the following question: what is the probability that a λ-sized likelihood region, drawn around a realization of X, will have had a prior chance less than c of containing μ?

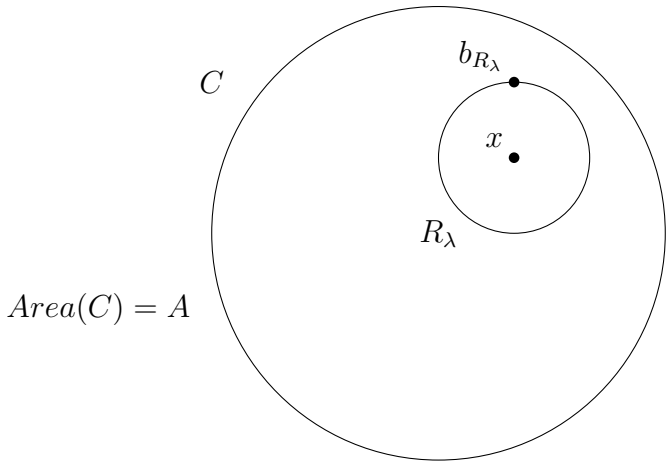


Figure A.1: Dartboard example, with likelihood region R_λ

For any fixed point x, the probability that X will be realized as x, where $R_\lambda(x)$ had less than a c chance of containing μ is just given by[1]

$$\begin{aligned}
&p(X = x \wedge p(\mu \in R_\lambda(x)) < c) \\
&= p(x|\mu \in R_\lambda(x)) \cdot p(\mu \in R_\lambda(x)|p(\mu \in R_\lambda(x)) < c) \\
&\qquad + p(x|\mu \notin R_\lambda(x)) \cdot p(\mu \notin R_\lambda(x)|p(\mu \in R_\lambda(x)) < c)) \\
&< \phi(x|\mu = x) \cdot c + \phi(x|\mu = b_{R_\lambda}) \cdot 1
\end{aligned}$$

But that upper bound, notice, is constant for every possible realization of X. So the overall probability that a λ-sized likelihood region – drawn around

[1]A bit of commentary about the proceeding calculation: Refer to Figure 1, and consider any point x on the dartboard. Now suppose that a λ-sized likelihood region around that point had a very small prior chance of containing μ – that is, of containing the point your friend was aiming at. Now ask: how would the world have to be arranged to make your friend's hitting x as probable as possible? There are two ways things could go here – μ could be in R_λ, even though it was antecedently improbable, in which case x is most probable if your friend is aiming directly at x. Alternatively, μ could fall outside R_λ, in which case your friend's aiming at a point directly on the boundary of R_λ makes x most probable.

the realization of your random variable X – will have had a prior chance less than c of containing μ is given by

$$\begin{aligned} p(X \wedge p(\mu \in R_\lambda(X)) < c) &= \int_C p(X = x \wedge p(\mu \in R_\lambda(x)) < c)\, dx \\ &< \int_C \phi(x|\mu = x) \cdot c + \phi(x|\mu = b_{R_\lambda})\, dx \\ &< \underbrace{\phi(x|\mu = x) \cdot c + \phi(x|\mu = b_{R_\lambda})}_{\text{Constant with respect to } x} \int_C 1\, dx \\ &< [\phi(x|\mu = x) \cdot c + \phi(x|\mu = b_{R_\lambda})] \cdot A \end{aligned}$$

Finally, notice that we can arbitrarily decrease the value of both summands in that last expression. $\phi(x|\mu = b_{R_\lambda})$ decreases with increases in our chosen λ (because the points on the boundary of R_λ get further from x as λ increases). And $\phi(x|\mu = x) \cdot c$ decreases with decreases in our choice of c. We can therefore come to know, prior to observing the realization of X, that it was arbitrarily improbable that we should witness a realization of X, where its associated λ-sized likelihood region had a prior probability less than c of containing μ. We simply have to be careful in our choices of c and λ.

References

Adam Koberinski, L. D., & Harper, W. L. (in press). Do the epr correlations pose a problem for causal decision theory? *Synthese*.

Adler, J. (2017). Epistemological problems of testimony. *The Stanford Encyclopedia of Philosophy*. Retrieved from `https://plato.stanford.edu/archives/win2017/entries/testimony-episprob/`

Ahmed, A., & Caulton, A. (2014). Causal decision theory and epr correlations. *Synthese*, *191*, 4315 - 4352.

Bertrand, J. (1889). *Calcul des probabilites*. Gauthier-Villars.

Bulmer, M. (1967). *Principles of statistics*. Dover Publications.

Carr, J. (in press). Accuracy or coherence? *Philosophy and Phenomenological Research*.

de Duve, C. (1991). *Blueprint for a cell: The nature and origin of life*. Neil Patterson.

Dempster, A. P. (1964). On the difficulties inherent in Fisher's fiducial argument. *Journal of the American Statistical Association*, *59*(305), 56–66.

Edwards, A. W. F. (1972). *Likelihood: an account of the statistical concept of likelihood and its application to scientific inference*. Cambridge University Press.

Eells, E. (1981, August). Causality, utility, and decision. *Synthese*, *48*(2), 295-239.

Egan, A. (2007). Some counterexamples to causal decision theory. *Philosophical Review*, *116*, 93 - 114.

Fisher, R. A. (1934). *Statistical methods for research workers*. Oliver and Boyd.

Fisher, R. A. (1935). Statistical tests. , *136*.

Fisher, R. A. (1956). *Statistical methods and scientific inference*. Oliver and Boyd.

Fry, I. (2000). *The emergence of life on earth*. Rutgers University Press.

Gandenberger, G. (2016, May). Why I am not a likelihoodist. *Philosopher's Imprint*, *16*(7).

Gibbard, A., & Harper, W. (1978 [1981]). Counterfactuals and two kinds of expected utility. In *Foundations and applications of decision theory* (p. 125 - 162). Dordrecht.

Greco, D. (2011). Significance testing in theory and practice. *British Journal for the Philosophy of Science*, *62*, 607-637.

Hacking, I. (1965). *Logic of statistical inference*. Cambridge University Press.

Hawking, S. (1996). *The illustrated a brief history of time*. Random House.

Horwich, P. (1982). *Probability and evidence*. Cambridge University Press.

Howson, C., & Urbach, P. (2006). *Scientific reasoning, the bayesian approach* (Third ed.). Carus Publishing Company.

Hoyle, F., & Wickramasinghe, C. (1981). *Evolution from space*. J.M. Dent and Sons.

Jaynes, E. T. (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, *3*(4), 227–241.

Jeffrey, R. (1983). *The logic of decision* (second ed.). University of Chicago Press.

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge University Press.

Knab, B. (2016). Origins of life research does not rest on a mistake. *Ergo*.

Krulwich, R. (2012). *Which is greater, the number of sand grains on earth or stars in the sky?* (Blog No. September 17). www.npr.org/sections/krulwich/2012/09/17/161096233/which-is-greater-the-number-of-sand-grains-on-earth-or-stars-in-the-sky. National Public Radio.

Lehmann, E. L. (1993). The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? *Journal of the American Statistical Association*, *88*(424), 1242–1249.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *231*, 289-337.

Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in honor of carl g. hempel* (p. 114-146). Reidel.

Popper, K. (1959). *The logic of scientific discovery*. Routledge.

Price, H. (1991). Agency and probabilistic causation. *British Journal of the Philosophy of Science*, *42*, 157-176.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*. Cambridge University Press.

Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge University Press.

Stalnaker, R. (1980 [1972]). Letter to david lewis. In *Ifs: Conditionals, belief, decision, chance and time* (p. 151 - 152). Springer.

van Fraassen, B. (1989). *Laws and symmetry*. Clarendon Press.

White, R. (2007). Does origins of life research rest on a mistake? *Nous*, *41*(3), 453–477.

Williamson, T. (n.d.). *Knowledge and its limits*. Oxford University Press.